

RESEARCH ARTICLE

Validation of Next-Generation Sequencing of Entire Mitochondrial Genomes and the Diversity of Mitochondrial DNA Mutations in Oral Squamous Cell Carcinoma

Anita Kloss-Brandstätter^{1*}, Hansi Weissensteiner^{1,2}, Gertraud Erhart¹, Georg Schäfer³, Lukas Forer^{1,2}, Sebastian Schönherr^{1,2}, Dominic Pacher^{1,2}, Christof Seifarth³, Andrea Stöckl¹, Liane Fendt¹, Irma Sottas³, Helmut Klocker³, Christian W. Huck⁴, Michael Rasse⁵, Florian Kronenberg¹, Frank R. Kloss⁵

1 Division of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria, **2** Department of Database and Information Systems, Institute of Computer Science, Leopold-Franzens University of Innsbruck, Innsbruck, Austria, **3** Department of Urology, Medical University of Innsbruck, Innsbruck, Austria, **4** Institute of Analytical Chemistry and Radiochemistry, Leopold-Franzens University of Innsbruck, Innsbruck, Austria, **5** Department for Cranio-, Maxillofacial and Oral Surgery, Medical University of Innsbruck, Innsbruck, Austria

* anita.kloss@i-med.ac.at



CrossMark
click for updates

OPEN ACCESS

Citation: Kloss-Brandstätter A, Weissensteiner H, Erhart G, Schäfer G, Forer L, Schönherr S, et al. (2015) Validation of Next-Generation Sequencing of Entire Mitochondrial Genomes and the Diversity of Mitochondrial DNA Mutations in Oral Squamous Cell Carcinoma. PLoS ONE 10(8): e0135643. doi:10.1371/journal.pone.0135643

Editor: Yong-Gang Yao, Kunming Institute of Zoology, Chinese Academy of Sciences, CHINA

Received: May 21, 2015

Accepted: July 23, 2015

Published: August 11, 2015

Copyright: © 2015 Kloss-Brandstätter et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data have been deposited to GenBank: <http://www.ncbi.nlm.nih.gov/genbank>. The accession numbers for the 35 reported individuals in this paper are KC286583 – KC286589 corresponding to Lab005 – Lab011 and KC286590 – KC286617 corresponding to MKG01 – MKG28.

Funding: The project was supported by grants of the Austrian Cancer Society/Tirol and the Tiroler Wissenschaftsfond, and Tiroler Standortagentur, which were awarded to A. Kloss-Brandstätter. The funders had no role in study design, data collection

Abstract

Background

Oral squamous cell carcinoma (OSCC) is mainly caused by smoking and alcohol abuse and shows a five-year survival rate of ~50%. We aimed to explore the variation of somatic mitochondrial DNA (mtDNA) mutations in primary oral tumors, recurrences and metastases.

Methods

We performed an in-depth validation of mtDNA next-generation sequencing (NGS) on an Illumina HiSeq 2500 platform for its application to cancer tissues, with the goal to detect low-level heteroplasmies and to avoid artifacts. Therefore we genotyped the mitochondrial genome (16.6 kb) from 85 tissue samples (tumors, recurrences, resection edges, metastases and blood) collected from 28 prospectively recruited OSCC patients applying both Sanger sequencing and high-coverage NGS (~35,000 reads per base).

Results

We observed a strong correlation between Sanger sequencing and NGS in estimating the mixture ratio of heteroplasmies ($r = 0.99$; $p < 0.001$). Non-synonymous heteroplasmic variants were enriched among cancerous tissues. The proportions of somatic and inherited variants in a given gene region were strongly correlated ($r = 0.85$; $p < 0.001$). Half of the patients shared mutations between benign and cancerous tissue samples. Low level heteroplasmies (<10%) were more frequent in benign samples compared to tumor samples, where

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

heteroplasmies >10% were predominant. Four out of six patients who developed a local tumor recurrence showed mutations in the recurrence that had also been observed in the primary tumor. Three out of five patients, who had tumor metastases in the lymph nodes of their necks, shared mtDNA mutations between primary tumors and lymph node metastases. The percentage of mutation heteroplasmy increased from the primary tumor to lymph node metastases.

Conclusions

We conclude that Sanger sequencing is valid for heteroplasmy quantification for heteroplasmies $\geq 10\%$ and that NGS is capable of reliably detecting and quantifying heteroplasmies down to the 1%-level. The finding of shared mutations between primary tumors, recurrences and metastasis indicates a clonal origin of malignant cells in oral cancer.

Introduction

Squamous cell carcinoma of the oral cavity (OSCC) account for 95% of all malignant lesions of the mouth and have become almost synonymous with oral cancer [1]. The development of OSCC is a multistep process modulated by genetic predisposition, chronic inflammation, tobacco and alcohol abuse and viral infections typically acquired via oral-genital contact [1, 2]. The overall five-year survival rate of OSCC is ~50% and has not changed significantly during the past 30 years [3, 4]. Therapeutic decisions are usually based on clinical and histopathological parameters, which however often fail to predict patient outcome and therapy success. Therefore, the search for new prognostic and predictive factors of OSCC reflects the need for improved risk assessment to customize therapeutic approaches.

Mitochondria have been implicated in the process of carcinogenesis because of their vital role in energy production [5], control of metabolic pathways [6] and the high mutation rate of their DNA (mtDNA) [7]. The most important aspect for medical mtDNA research is the occurrence of heteroplasmy, which is defined as the coexistence of two or more populations of mtDNA molecules with slightly different nucleotide composition in a single mitochondrion, cell, tissue or individual. Heteroplasmic mtDNA mutations were found in normal human cells [8–11] as well as in many human tumors [12–15].

Previously, our group could demonstrate that prostate tumor-specific mtDNA heteroplasmies were associated with elevated PSA-levels and might thus allow early detection of prostate cancer [14]. Recent studies on the involvement of mtDNA mutation in OSCC suggest that pathogenic mtDNA heteroplasmies are potential prognostic markers for OSCC [16]. The detection of low levels of heteroplasmy in tumors may allow unprecedented early identification and monitoring of neoplastic progression to malignancy. Therefore, a sensitive detection and accurate quantification of heteroplasmic changes may assist in a personalized risk assessment.

Conventional Sanger sequencing is not sensitive enough to detect low level heteroplasmy, i.e. heteroplasmy <10%. In addition, the ratio of peak heights at heterozygous positions or any other sequence variant are context-dependent and may differ strongly between forward and reverse strands [17]. However, NGS on an Illumina HiSeq offers a robust platform for comprehensive mtDNA analysis, as the small size of the mitochondrial genome and the resulting high coverage for each nucleotide position enable a more sensitive and accurate quantification of low-level heteroplasmy [18]. Currently, we are in the midst of a transition process from Sanger

sequencing to NGS-based molecular diagnosis of mtDNA disorders, and several studies have already demonstrated that NGS on an Illumina platform is sensitive and accurate in detecting heteroplasmic mtDNA changes [15, 19–22].

The purpose of this research project was to evaluate the application of next-generation sequencing on an Illumina HiSeq platform for detecting and quantifying heteroplasmic mixtures in cancer tissue samples. The second goal was to explore the variation of somatic mtDNA mutations in primary oral carcinomas, recurrences and metastases.

Twenty-eight oral cancer patients were prospectively enrolled and followed for a median of 30.5 months. The entire mitochondrial genome of 85 tissue samples (primary tumors, recurrences, resection margins, blood, lymph node metastases, one dysplasia and one second primary tumor) was genotyped in 91% of cases with both Sanger and next-generation sequencing, except for a few samples, where not enough DNA was available for both sequencing procedures. The combined evaluation of clinical findings, histopathological analyses and of the results from the genetic analyses of different tissues enabled an assessment of the efficiency of low-level mtDNA mutation screening.

Results

Determination of the lower detection limit of heteroplasmies

In order to determine the lower detection limit of heteroplasmies, four samples with pre-defined mixtures of the DNA of two lab technicians (Lab002 and Lab011) with mixture ratios of 1+1, 1+9, 1+49 and 1+99 were subjected to both Sanger [14] and next-generation sequencing. The two technicians differed on 27 nucleotide positions dispersed over the entire genome. Despite the high sequence coverage and the excellent signal-to-noise ratios in the sequence electropherograms from Sanger sequencing, only heteroplasmic mixtures at the 7% level could be detected (Fig 1). For NGS, the samples were sequenced on an Illumina HiSeq 2500 platform. With the high NGS coverage for the four mixture sample as well as the two original DNA samples, we were able to exactly quantify the minor components at the 1% level as obtained from the 1+99 mixture by lowering the detection threshold to 0.8% within the mtDNA-Server application. Position 16189 was not found at the 1% level. However, we found three additional mutations at 9462, 15236 and 16129, which were confirmed as private heteroplasmic mutations from the original DNA sample Lab01. These mutations however were true and cannot be regarded as false positives, as they were detected in the original DNA sample and in the 1+1-mixture. All results from the mixture experiment are given in S4 Table.

Comparison between Sanger and next-generation sequencing

Similar to a study on nuclear DNA mutations [23], the limit of detection for Sanger sequencing was identified at 7% of mutant mtDNA to the reference genome rCRS [24]. As all samples from our study were subjected to both Sanger and NGS on an Illumina platform, we compared the estimates of the percentages of the minor components in point heteroplasmies between Sanger electropherograms (as calculated with Sequencher) and NGS reads for all point heteroplasmies and found a very strong correlation between the two measurements (Pearson's $r = 0.99$; $p < 0.001$). For heteroplasmies $< 10\%$, which were detected with NGS, but could not be found in Sanger electropherograms, because they could not be distinguished from the sequence background, the value for the Sanger estimation was set to $\frac{10}{\sqrt{2}}$, corresponding to *Limit of detection* / $\sqrt{2}$ as suggested Croghan and Egehy [25].

However, a high correlation for any two methods designed to measure the same property does not automatically imply that there is good agreement between the two methods [26].

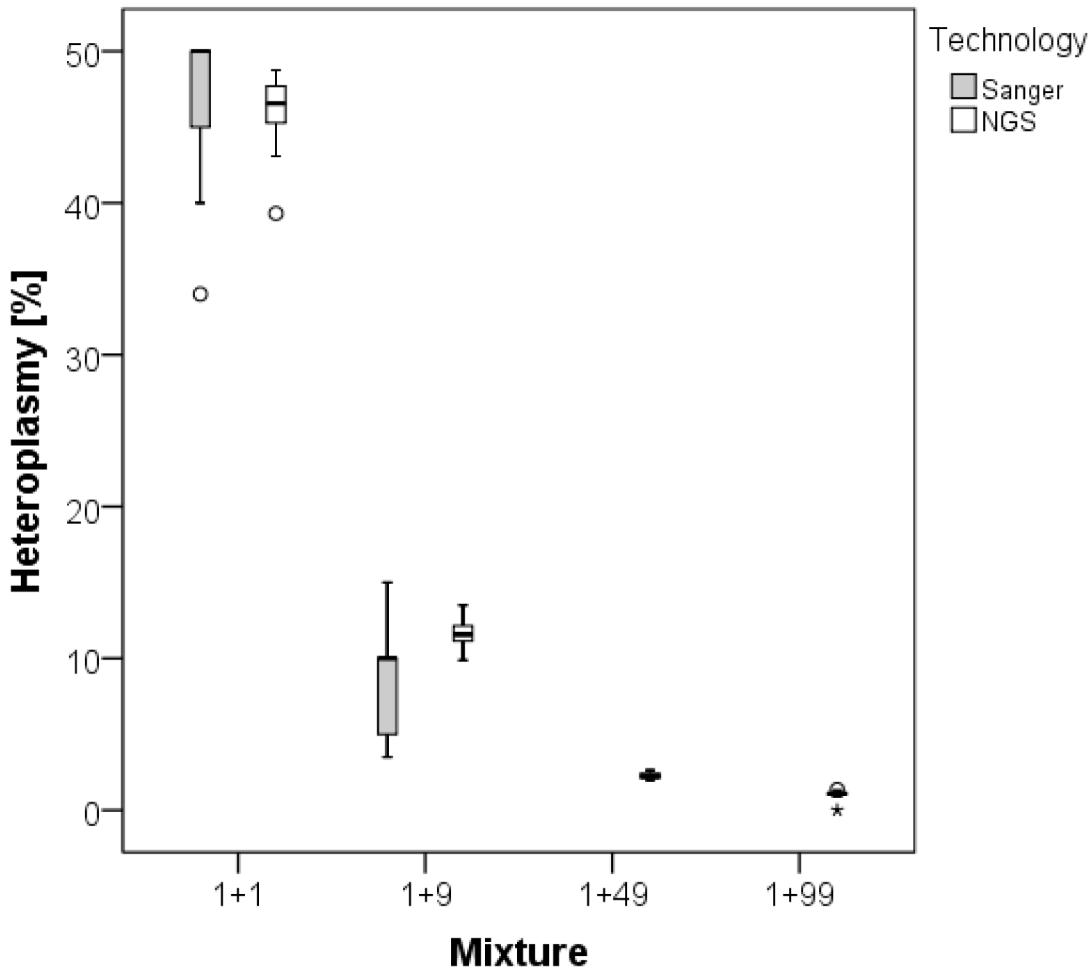


Fig 1. Percentage of the minor component on heteroplasmic positions as obtained with Sanger sequencing and next-generation sequencing on an Illumina HiSeq platform. The samples contained pre-defined mixtures of the DNA of two lab members. In Sanger electropherograms, heteroplasms were neither detected in the 1+49 mixture nor in the 1+99 mixture.

doi:10.1371/journal.pone.0135643.g001

Therefore, a Bland-Altman plot was created for analyzing the agreement between Sanger sequencing and NGS heteroplasmy measurements (Fig 2). All but ten values lay within the 95% limits of agreement. The mean value of the difference (NGS–Sanger = -1.62) between the two measurements was significantly different from 0 ($p < 0.001$), indicating that there was a systematic difference between the two methods, with Sanger sequencing slightly underestimating the heteroplasmy level. However, as that 95.3% of values lay within the 95% limits of agreement, we concluded that NGS and Sanger sequencing can be used interchangeably for determining the mixture ratios on point heteroplasmic positions, where the minor component amounted for at least 10% of the nucleotide mixture.

Mitochondrial DNA sequences

The entire mitochondrial genome was sequenced using both Sanger and Illumina HiSeq next-generation sequencing from 77 of the 85 tissue samples (primary tumors, recurrences, resection margins of both primary tumors and recurrences, blood, cancerous and benign areas in lymph nodes, one dysplasia and one second primary tumor) (S1 Table) collected from 28 prospectively enrolled oral cancer patients (Fig 3). Only from 8 samples (6 paraffin samples, two

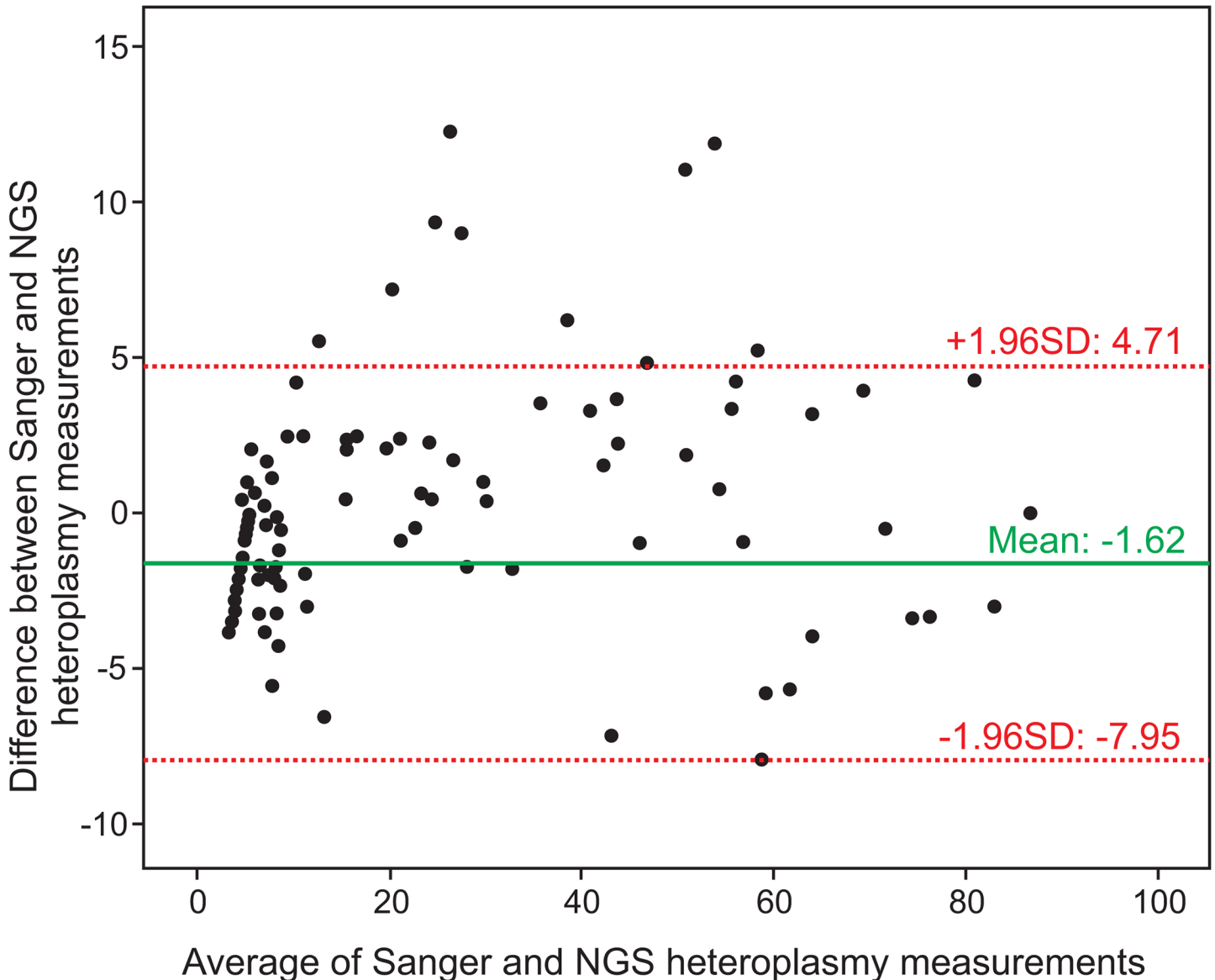


Fig 2. Bland-Altman plot for depicting the agreement between NGS and Sanger sequencing heteroplasmy measurements. On the x-axis, the average of next-generation and Sanger sequencing heteroplasmy estimates is plotted. On the y-axis, the difference between next-generation and Sanger sequencing heteroplasmy estimates is plotted. The mean difference is indicated as green line, the 95% limits of agreement (average difference \pm 1.96 standard deviation of the difference) are indicated as red dotted lines. For heteroplasmy 10%, which were detected with NGS, but could not be found in Sanger electropherograms, the value for the Sanger estimation was set to $\frac{10}{\sqrt{2}}$.

doi:10.1371/journal.pone.0135643.g002

benign tissue samples) we could not extract enough DNA to subject them to both sequencing technologies, so they were only typed with Sanger sequencing (S2 Table). The sequences from benign tissue samples were stored in GenBank (accession numbers KC286583 –KC286589 corresponding to Lab005 –Lab011 and KC286590 –KC286617 corresponding to MKG01 –MKG28).

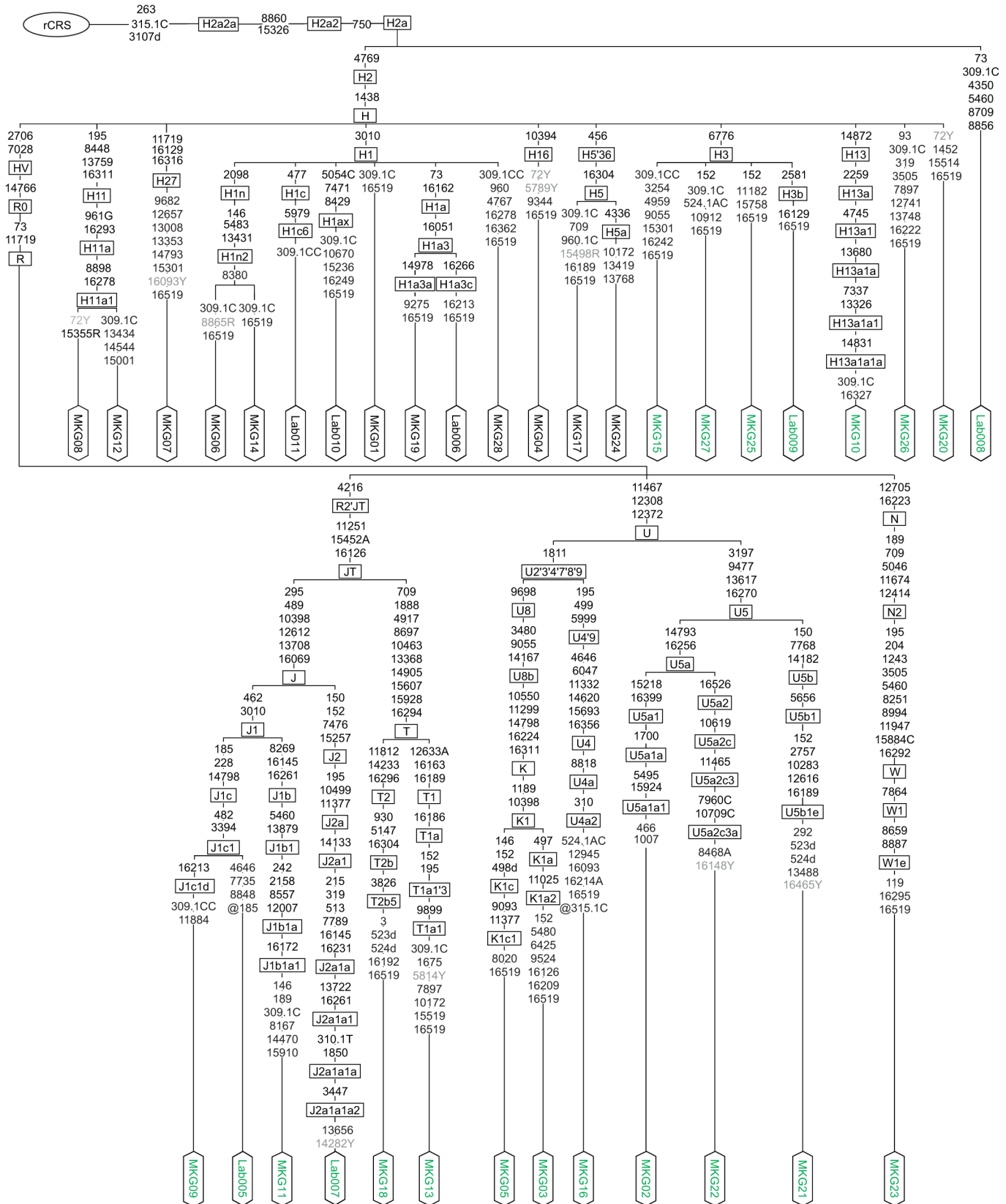


Fig 3. Phylogenetic tree representing the benign profiles of 28 oral cancer patients and seven team members. The sequences are stored in GenBank (accession numbers KC286583–KC286589 corresponding to Lab005–Lab011 and KC286590–KC286617 corresponding to MKG01–MKG28). The position of the rCRS is indicated for reading off sequence motifs. Mutations are transitions unless a base is explicitly indicated. The prefix “@” designates reversions, whereas suffixes indicate transversions (to A, G, C, or T), indels (.1, d), and heteroplasmies (R, Y). Point heteroplasmic mutations that were observed in benign tissue samples are highlighted in grey.

doi:10.1371/journal.pone.0135643.g003

Distribution of inherited and somatic variations across genomic regions

All observed point and length heteroplasmic mutations are described in [S2 Table](#). Length heteroplasmy occurred frequently within the polycytosine stretches of HVS-I (anywhere between positions 16188 and 16195) and HVS-II (around position 310) due to replication slippage when a specific minimum number of cytosine residues are present [8, 14]. Interestingly, tumor samples showed a higher tendency to exhibit length heteroplasmy than their corresponding benign tissue samples. Four tumor samples (MKG01, MKG10, MKG12 and MKG26) showed additional cytosine-insertions in the HVS-II C-stretch after position 309 compared to their benign tissue samples, while only one tumor sample (MKG11) showed fewer C-residues in the HVS-II C-stretch with respect to its benign tissue sample. One tumor sample (MKG13) showed length heteroplasmy after position 573 in HVS-III, but its corresponding benign sample did not show any signs of heteroplasmy at this position. The most interesting observation was that two unrelated tumor samples (MKG11 and MKG21) exhibited frame-shift causing length heteroplasmy in MT-ND5 on position 12390, which was not seen in the associated benign samples.

All OSCC patients showed somatic mtDNA alterations in their index tumors and/or benign tissue samples. However, when considering Sanger sequencing data only, then tumor-specific mtDNA heteroplasmies could only be detected in 20 out of 28 patients. We observed cancer-specific variants (primary tumors, recurrences and lymph node metastasis) on 124 different nucleotide positions. Homoplasmic polymorphisms that were detected in all tissue samples of a patient were regarded as germline variants. Most somatic variations were singletons, arising in only one patient. In total, 6.9% of somatic mutations were within tRNAs, 18.3% within ribosomal RNAs, 52.2% in protein-coding regions, and 22.6% in the control region.

There was a strong correlation between the proportion of somatic variants and the proportion of inherited variants in a given gene region ($r = 0.85$; $p = 3.7 \times 10^{-5}$), and among protein-coding genes, both the proportion of somatic ($r = 0.81$; $p = 0.001$) and inherited nucleotide variations ($r = 0.74$; $p = 0.005$) correlated with the proportional length of the gene in relation to the entire mitochondrial genome ([S1 Fig](#)). In addition, there was a significant enrichment of non-synonymous variants among cancerous tissue samples as compared with germline variants ($p < 0.001$).

Comparison with other NGS-typed oral cancer mitochondrial genomes

We were interested whether the distribution of heteroplasmies across genetic regions as observed in our study could be characteristic for other oral cancer mitochondrial genomes, as this would corroborate the authenticity of the variations that we found. Therefore, we compared the frequency distribution of heteroplasmic sites found in our tumor samples with the frequency distribution of heteroplasmies detected in 47 head and neck tumor samples and with 1907 samples of diverse carcinomas [27] derived from whole genome and whole exome sequencing. Since Ju et al. applied a 3% filter for the variant allele frequency, we restricted our data specifically for this comparative analysis to heteroplasmies $> 3\%$. The dispersion of heteroplasmies in our sample was highly correlated with the heteroplasmy distribution in head and neck tumors ($r = 0.79$; $p < 0.001$) and with the heteroplasmy allocation in diverse tumor

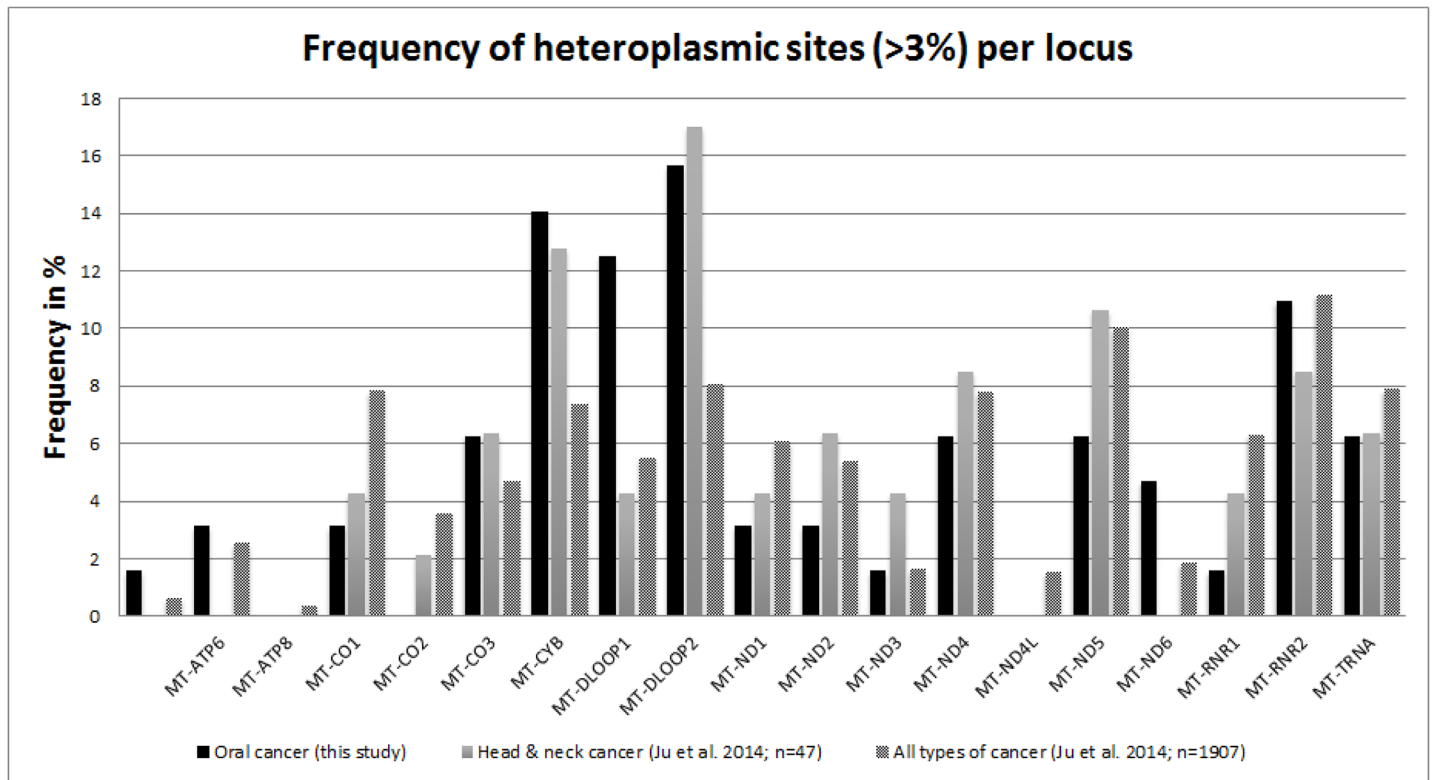


Fig 4. Distribution of NGS-typed heteroplasmies across the mitochondrial genome. Frequencies were calculated as number of heteroplasmies per gene region divided by the total number of observed heteroplasmies. The data from our study were compared with data from head and neck cancers and with all cancer samples from Ju et al. [27].

doi:10.1371/journal.pone.0135643.g004

samples ($r = 0.62$; $p < 0.001$) (Fig 4). Therefore, the frequency of heteroplasmies in diverse gene regions found in our study seems to reflect well the distinct vulnerability of diverse areas of the mitochondrial genome to cancerous somatic mutations.

Heteroplasmic mutations shared between benign and cancerous tissue samples

One recent study suggested that many putative somatic mutations are in fact low-level heteroplasmies undetected in the germline that have undergone clonal expansion in the tumor [28]. In our study, 54% of the patients shared mutations between benign and cancerous tissue samples (MKG01: m.G14560R; MKG04: m.T5789Y; MKG05: m.C16261Y; MKG06: m.G8865R; MKG07: m.T16093Y; MKG08: m.G15355R; MKG11: m.T146Y; MKG13: m.T5814Y and m.C13287Y; MKG15: m.A9794R; MKG17: m.G15498A; MKG21: m.G13759R and m.C16465Y; MKG22: m.C16148Y; MKG23: m.T4597Y; MKG27: m.A215R, m.A9702R and m.A16241R; MKG28: m.C64Y).

In MKG04 (m.T5789Y), MKG06 (m.G8865R) and MKG08 (m.G15355R), the heteroplasmies were also found in the blood samples of the patients, underscoring that those were in fact germline mutations. However, we only analyzed blood samples from 4 out of 28 patients (14.3%), and in those, we found consistencies between tumor and blood heteroplasmies in 3 out of 4 patients (75%).

Heteroplasmic mutations shared between different areas of the same tissue sample

For three patients, distinct areas within the same tumor entity were used for microdissection, because these areas differed from each other from a histopathologically perspective. In patient MKG01, two separate cancerous areas were sequenced, which both shared exactly the same three mutations (m.G14560A, m.T15033C and m.G15553A). This observation was in contrast to patient MKG03, where two separate tumor areas showed a completely different mutation profile (primary tumor area 1: m.G5881A, m.T11790C, m.T14787C and m.G15106A; primary tumor area 2: m.T10983C). For patient MKG27, two distinct benign areas were sampled, which shared one heteroplasmic mutation (m.A16241G), but also differed on four other positions (benign area 1: m.A3523G and m.A14002G; benign area 2: m.A215G and m.A9702G). In addition, three distinct tumor areas were sequenced, which shared only one mutation (m.A215G). Primary tumor area 2 shared another mutation with primary tumor area 1 (m.G2690A), and primary tumor area 1 shared four mutations with primary tumor area 3 (m.A183G, m.A9702G, m.T13897C and m.A16241G), which showed on mutation that was not observed in any other benign or cancerous area (m.A1082G). To sum up, in samples where different areas within the same tumor were selected for DNA extraction and mtDNA sequencing, we observed both concordances and discordances in the mutational pattern of mtDNA heteroplasmies.

Somatic mutations in recurrences

Six patients developed a tumor recurrence (MKG05, MKG10, MKG13, MKG21, MKG24 and MKG27) during follow-up, four of these patients (MKG05, MKG10, MKG13 and MKG21) showed mutations in the recurrence that had also been observed in the index tumor and in three of these patients (MKG05, MKG13 and MKG21), the mutations were also observed in the resection margins (Table 1). On three out of five nucleotide positions, patient MKG05 showed an increase of the proportions of the mutated nucleotides from the index tumor (~13%) to the recurrence (~70%). However, on two nucleotide positions in the mtDNA profile

Table 1. Occurrence of somatic mtDNA mutations in different tissues of the same patient.

Patient	Mutation	Benign	Blood	Primary tumor	Recurrence	Tumor in lymph node
MKG05	C64Y	n.d.	n.d.	12.13	62.3	n.a.
MKG05	A5894R	n.d.	n.d.	15.54	71.6	n.a.
MKG05	G6762R	n.d.	n.d.	7.14	1.2	n.a.
MKG05	G10310R	n.d.	n.d.	10.8	61.2	n.a.
MKG10	T9865Y	n.d.	n.a.	1.76	1.6	n.a.
MKG15	C5297Y	n.d.	n.a.	45.87	n.a.	84.8
MKG15	G9565R	n.d.	n.a.	12.41	n.a.	55.1
MKG15	G12868R	n.d.	n.a.	37.8	n.a.	85.8
MKG20	G2916R	n.d.	n.a.	21.02	n.a.	31.5
MKG20	G12736R	n.d.	n.a.	24.73	n.a.	32
MKG27	A183R	n.d.	n.a.	22.7	n.a.	57.5
MKG27	T13897Y	n.d.	n.a.	2.4	n.a.	1.9

Notes: The tables indicate the percentage of the mutated nucleotide (relative to the rCRS) on point-heteroplasmic positions. Only mutations, which occurred in more than one tissue sample of a patient, were included in these tables.

n.a. tissue was not available

n.d. respective heteroplasmy was not detected above the threshold of 1%

doi:10.1371/journal.pone.0135643.t001

Table 2. Occurrence of shared mtDNA mutations considered “germline mutations”

Patient	Mutation	Benign	Blood	Primary tumor	Recurrence	Tumor in lymph node
MKG01	G14560R	1.65	n.a.	1.38	n.a.	n.a.
MKG04	T5789Y	16.91	45.62	8.21	n.a.	n.a.
MKG05	C16261Y	1.27	n.d.	87.02	74.8	n.a.
MKG06	G8865R	27.74	20.86	65.87	n.a.	n.a.
MKG07	T16093Y	22.39	n.a.	10.58	n.a.	n.a.
MKG08	G15355R	52.08	56.66	10.37	n.a.	n.a.
MKG11	T146Y	98.73	45.16	n.a.	n.a.	n.a.
MKG13	C13287Y	3.53	n.a.	1.38	1.7	n.a.
MKG13	T5814Y	6.73	n.a.	4.15	5.3	n.a.
MKG15	A9794R	2.84	n.a.	1.63	n.a.	74.7
MKG17	G15498R	23.73	n.a.	30.5	n.a.	n.a.
MKG21	G13759R	2.66	n.a.	1.26	1.54	n.a.
MKG21	C16465Y	89.32	n.a.	41.85	43.35	n.a.
MKG22	C16148Y	30.5	n.a.	6.17	n.a.	n.a.
MKG23	T4597Y	1.28	n.a.	1.3	n.a.	n.a.
MKG27	A16241R	1.1	n.a.	1.14	n.a.	n.d.
MKG27	A9702R	8.6	n.a.	25.4	n.a.	58.9
MKG27	A215R	24	n.a.	32.5	n.a.	56.7
MKG28	C64Y	6.83	n.a.	3.33	n.a.	n.a.

Notes: The tables indicate the percentage of the mutated nucleotide (relative to the rCRS) on point-heteroplasmic positions. Only mutations, which occurred in more than one tissue sample of a patient, were included in these tables. The recurrence in patient MKG21 should be regarded as second primary tumor, as the primary oral squamous cell carcinoma (OSCC) of this patient occurred in her mandible, while the second OSCC was found in her maxilla.

n.a. tissue was not available

n.d. respective heteroplasmy was not detected above the threshold of 1%

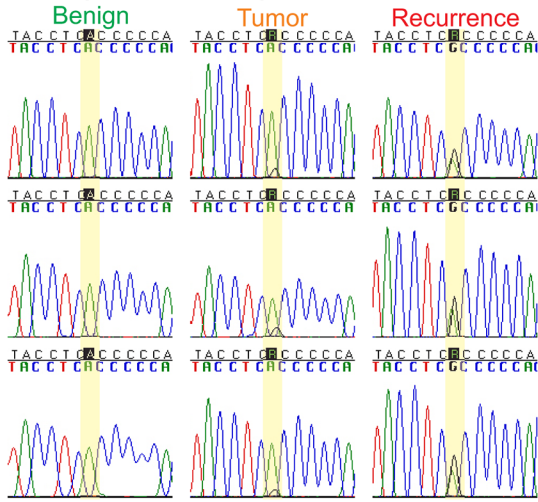
doi:10.1371/journal.pone.0135643.t002

of patient MKG05, we observed an average decrease of ~9% from index tumor heteroplasmies to recurrence heteroplasmies (Tables 1 and 2 and Fig 5). Patient MKG21 was first surgically treated because of OSCC in the mandible, and 16 months later a second OSCC was removed from this patient’s maxilla. The most intriguing observation was that the mandibular tumor showed length heteroplasmy after position 12390, which was also found in the maxillary tumor, but not in the benign area surrounding the first tumor (Fig 6 and S2 Table).

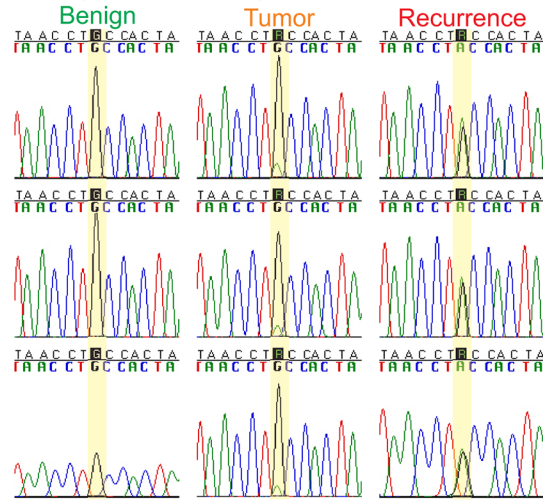
Patient MKG27 developed a recurrence, but died before a potential second surgery. Therefore, we did not have a tissue sample from the recurrence available for analysis. However, similar to patients MKG05, MKG13 and MKG21, we found two of the somatic tumor mutations also in the benign tissue surrounding the index tumor of patient MKG27 (Table 2). Eleven further patients (MKG01, MKG04, MKG06, MKG07, MKG08, MKG11, MKG15, MKG17, MKG22, MKG23 and MKG28) shared mutations between tumor areas and benign areas surrounding the tumor (S2 Table), but did not develop a recurrence during follow-up.

Those somatic mutations, i.e. mutations that were not present in the benign tissue, which were also found in either the recurrence or the lymph node metastases, amounted for on average 17.9% of the mixture, and increased to an average of 39.6% in recurrences and to an average of 49.8% in lymph node metastases (Table 1). The average heteroplasmy level as computed from S6 Table was 16.4%. Therefore, we saw no difference to the mean value of heteroplasmies in cancerous samples from the entire data set. However, it was interesting to see that 75% of somatic mutations, which were shared between tissues, were high-level (i.e. larger 10%), while

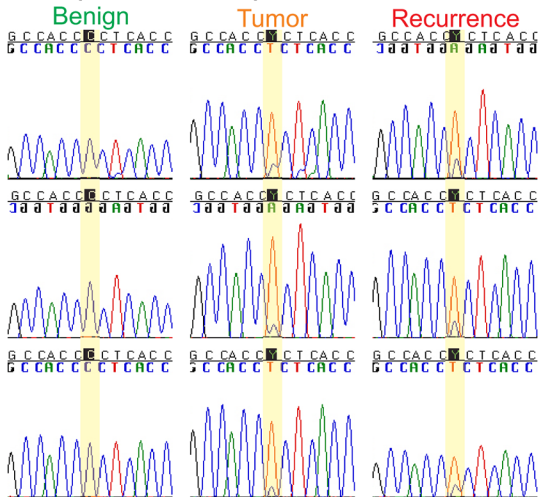
Sample MKG05, position m.A5894G



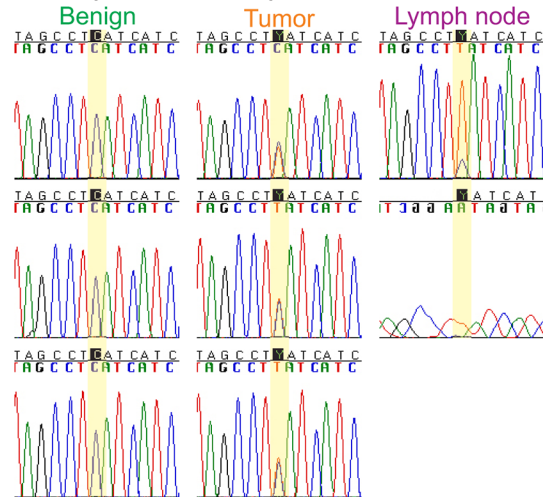
Sample MKG05, position m.G10310A



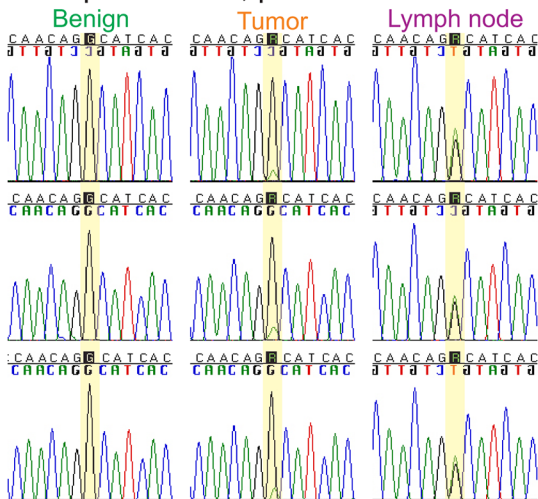
Sample MKG05, position m.C16261T



Sample MKG15, position m.C5297T



Sample MKG15, position m.G9565A



Sample MKG15, position m.G12868A

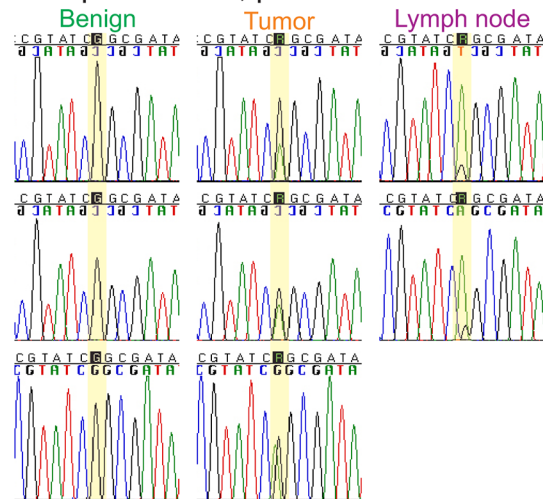


Fig 5. Heteroplasmic mutations as seen with Sanger sequencing in MKG05 and MKG15. The benign tissue was taken from the resection border of the tumor.

doi:10.1371/journal.pone.0135643.g005

in [S6 Table](#), only 36.1% of somatic mutations in cancer tissues were high-level. The observation that higher level heteroplasmies were more likely to be shared could be yield further evidence for the clone expansion hypothesis.

In summary, 83% of the recurrences shared heteroplasmic mutations with the primary tumor, and 50% of the recurrences also shared heteroplasmic mutations with the resection margins of the primary tumor. However, the sharing of heteroplasmic mutations between primary tumors and resection margins was not associated with tumor relapse in our study ($p > 0.05$).

Somatic mutations in lymph nodes

Five patients showed tumor metastases in the lymph nodes of their necks (MKG15, MKG18, MKG20, MKG26 and MKG27). Three of these patients (MKG15, MKG20 and MKG27) shared mtDNA mutations between index tumors and lymph node metastases ([Table 1](#)). In MKG26, the primary tumor showed only one somatic mutation, which was not found in the lymph node metastasis, which itself harbored six somatic mtDNA heteroplasmies. A consistent increase of the proportions of mutated nucleotides on all but one shared heteroplasmic sites from the oral carcinoma to the lymph node metastases was observed in all three patients (MKG15: from ~25% to ~70%; MKG20: from ~23% to ~32%; MKG27: from ~10% to ~44%) ([Figs 6 and 7](#)). This corresponded to an increase by a factor of 2.8 in MKG15, by a factor of 1.4 in MKG20, and by a factor of 4.2 in MKG27. The patients died rapidly (7–15 months) after tumor resection.

Comparison of heteroplasmy between tumor and benign tissue samples

When comparing heteroplasmy levels between tumor and benign tissue samples, it was obvious that low level heteroplasmies, defined as heteroplasmies $< 10\%$, were more frequent in benign compared to cancerous tissue samples, while heteroplasmies with minor alleles above 10% dominated in tumor samples ($p < 0.001$; [Fig 7](#)). Indeed, 65 out of 82 (79.3%) heteroplasmic mutations found in benign tissues were below 10% ([S5 Table](#)), while 62 out of 97 (63.9%) of heteroplasmic mutation found in tumor tissues were below 10% ([S6 Table](#); Chi-Square-Test: $p = 0.024$). However, when comparing the phylogenetic stability of the mutations between the different tissue types, there was no statistical evidence that the average phylogenetic weights as estimated with HaploGrep [[29](#)] or EMMA [[30](#)] differed between the four main tissue groups (benign, primary tumor, recurrence, metastases) ($p > 0.05$). Therefore, we found no evidence for a preference of mutating phylogenetically stable and thus probably physiologically important sites in tumor samples. However, given our sample size, we also lack statistical power for such a comparison.

Exclusion of potential next-generation sequencing artifacts

In order to exclude phantom mutations, which are systematic artifacts generated in the course of the sequencing process, the frequencies of heteroplasmic sites were inspected closely. Two positions, namely 414K and 72Y showed up in 12 and 11 different samples, respectively. It is suspicious that both positions share the same sequence neighborhood, namely GG[T]ATGC AC. This is similar to a previous description of phantom mutation hotspots in Sanger electropherograms, where a certain nucleotide motif was found to be responsible for most artifacts in

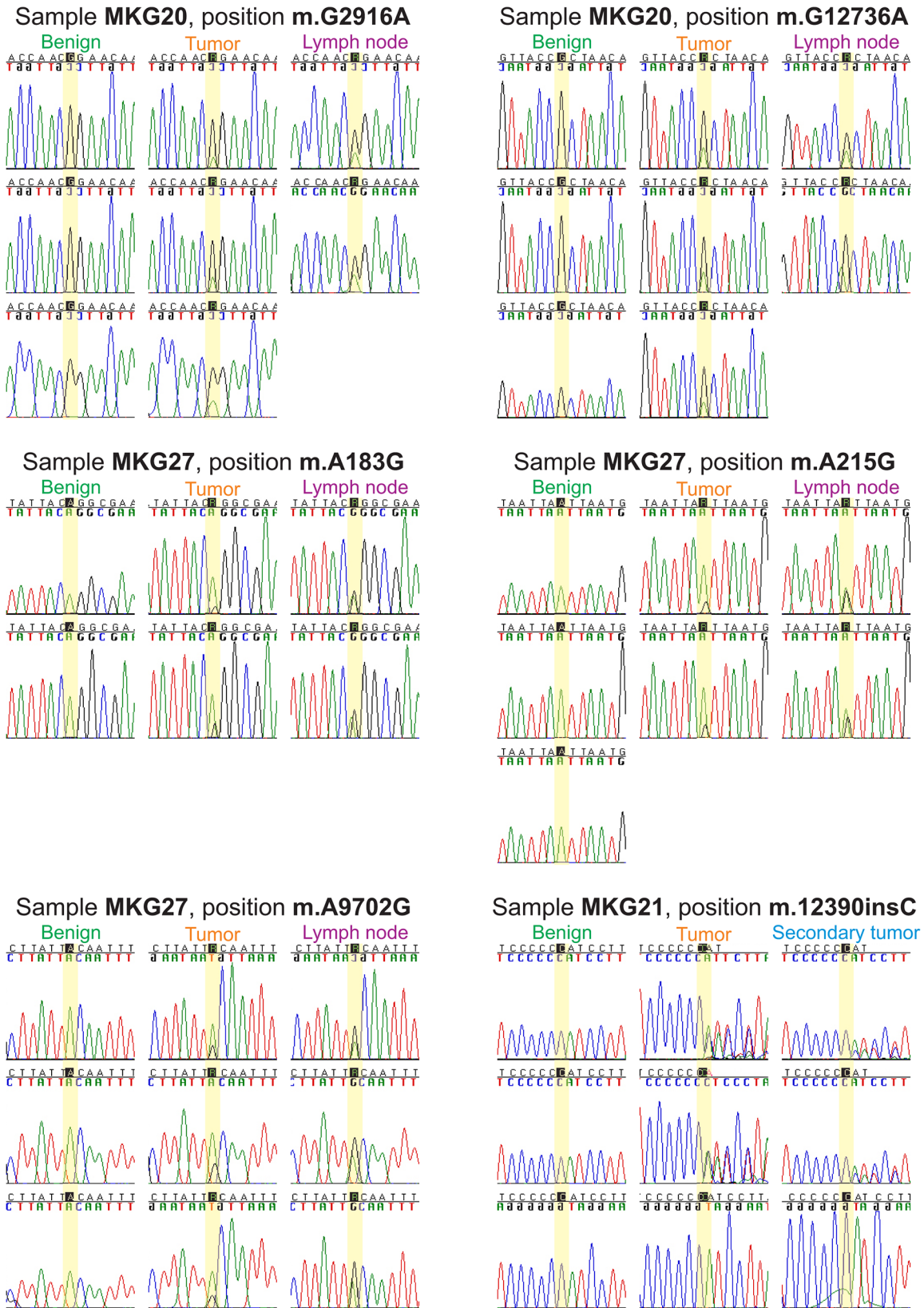


Fig 6. Heteroplasmic mutations as seen with Sanger sequencing in MKG20, MKG21 and MKG27. The benign tissue was taken from the resection border of the tumor.

doi:10.1371/journal.pone.0135643.g006

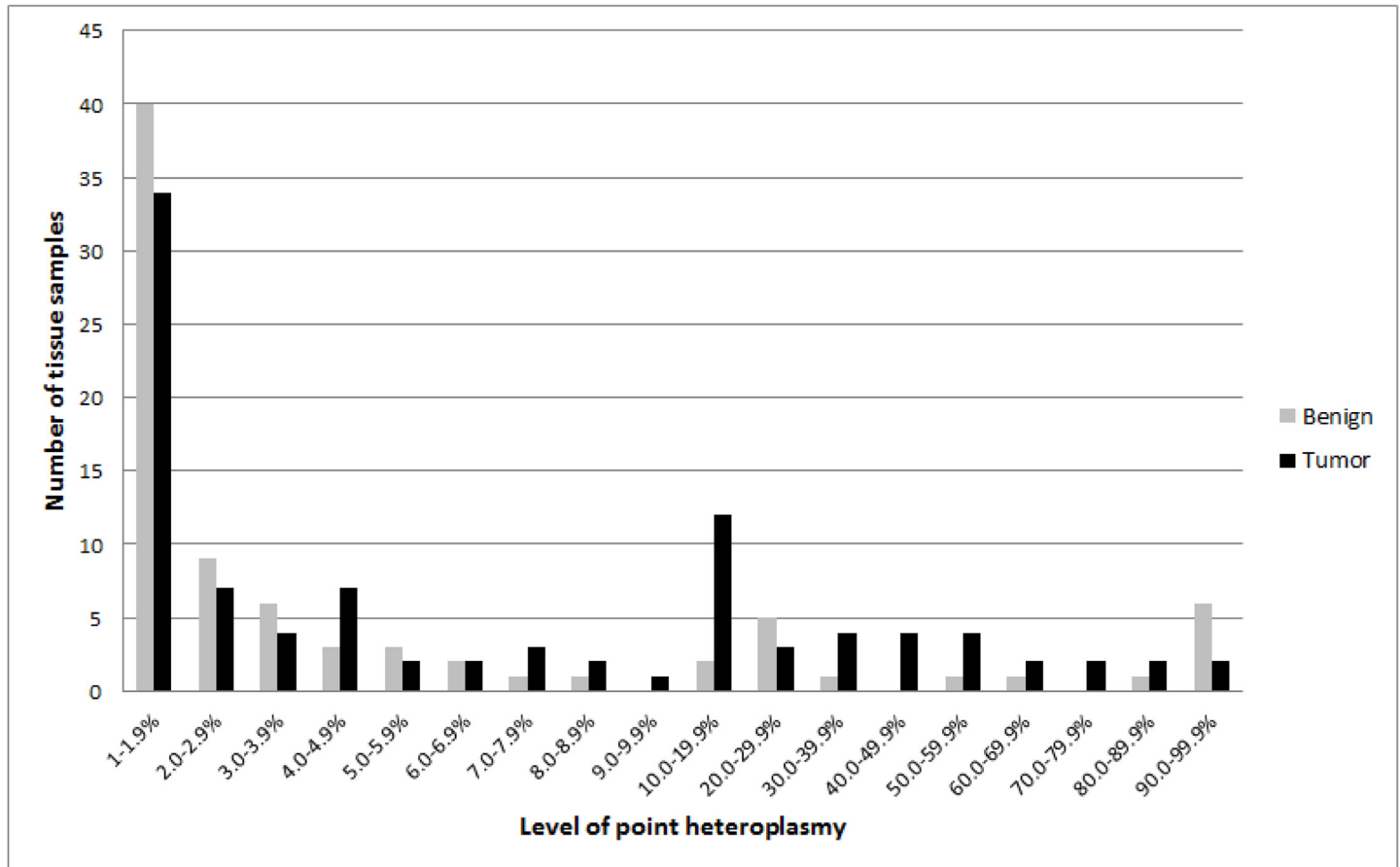


Fig 7. Distribution of point heteroplasmy levels in tumor and benign tissue samples.

doi:10.1371/journal.pone.0135643.g007

single-stranded sequences [31]. In addition, the coverage on the forward reads fell compared to the neighboring bases to ~52% on 72Y and ~73% on 414K, independently if BAQ [32] was applied or not. The reverse reads did not show any reduction, but a constant coverage at the specific positions. Furthermore, some samples showed 72K on the forward reads, but 72Y on the reverse reads. We hypothesized that low-level (<5%) heteroplasmy on these positions, which were not seen in Sanger electropherograms, could be considered as potential phantom mutations. This should however be verified for example with cloning and mini-sequencing [33] or droplet digital PCR. In addition, a recent study by Li and colleagues also observed a similar pattern on position 72 [34]. Unfortunately, while many other mutations were validated with droplet digital PCR in the study by Li et al., the 72Y was not further confirmed. Still, three samples from our study showed high levels of heteroplasmy at position 72 (MKG04 benign: 60.18%, MKG08 benign: 16.62% and MKG20 benign: 27.41%), that could be verified with Sanger sequencing. We therefore excluded eight (seven benign and one tumor sample) of the low-level heteroplasmy sites at 72Y, as well as 12 occurrences of T414K, which showed a mean heteroplasmy level of 1.5%. Samples with higher heteroplasmy levels at 72Y (MKG4 benign, MKG8 benign and MKG20 benign) all showed 73R and some of these samples also showed 74K (MKG4 benign and MKG20 benign), which levels were directly correlated with the occurrence of 72Y (S7 Table). We therefore excluded 73R and 74K. We further deleted apparent mutations surrounding long-range PCR primer annealing sites around position 2,600, where

we found unusual patterns. This was the case in samples MKG1 benign (2523–2667), MKG1 cancer (nps 2523–2667) and MKG5 benign (nps 2523–2667; please see [S3 Table](#) for unfiltered results). We also excluded site 3Y found in Sample 18, which is a homoplasmic mutation according Sanger Sequencing and could be regarded as the result of mapping problems, due to the circular nature of the mitochondrial genome.

Discussion

To the best of our knowledge, our study is the first investigation that provides a profound comparison of Sanger and next-generation sequencing technologies applied to entire mitochondrial genomes in cancer tissue samples. All samples were subjected to both high-quality Sanger sequencing (96 600bp-electropherograms per genome) and high-coverage NGS on an Illumina HiSeq platform (~35,000 reads per base). We observed a very strong correlation between the two methods in estimating the exact mixture ratio (Pearson's $r = 0.99$; $p < 0.001$). However, the Bland-Altman plot indicated that there was a small systematic difference between the measurements, with Sanger sequencing slightly underestimating the heteroplasmy levels. Still, as 95.3% of samples lay within the 95% limits of agreement, we conclude that the two methods can be used almost interchangeably for determining the mixture ratio on point heteroplasmic positions, where the minor component amounts for at least 10% of the nucleotide mixture. Nonetheless, it has to be said that the majority of heteroplasmies lay below the detection limit of Sanger sequencing. In addition, the detection and interpretation of point heteroplasmies in Sanger electropherograms strongly depends on the experience and expertise of the evaluating mtDNA scientist, and could easily vary between different evaluations. On the contrary, the procedure of heteroplasmy detection and quantification with NGS is completely independent from the evaluator and therefore an impartial and replicable process. The reliable detection of length heteroplasmies can be considered as strength of Sanger sequencing, however we did not evaluate NGS data regarding length heteroplasmies. The major advantage of Illumina HiSeq NGS is the exact quantification of nucleotide mixtures down to the 1%-level, an observation that is in agreement with recent data [\[35\]](#).

When taking all available information on point heteroplasmies in our dataset into consideration (including low level variants down to 1% nucleotide mixture ratio), we found a strong correlation between the proportion of somatic variants and the proportion of inherited variants in a given gene region ($r = 0.85$; $p = 3.7 \times 10^{-5}$), indicating that certain regions of the mitochondrial genome are more susceptible to mutations in general compared to other regions. However, this observation could also be explained by the fact that among protein-coding genes, both the proportional number of somatic ($r = 0.81$; $p = 0.001$) and inherited nucleotide variations ($r = 0.74$; $p = 0.005$) correlated with the proportional length of the gene, a finding that has also been described for breast cancer [\[36\]](#). Another interesting observation was that there was a significant enrichment of non-synonymous variants among cancerous tissue samples as compared with germline variants ($p < 0.001$). This finding is in line with a recent observation of an accumulation of somatic missense mutations in cancer mitochondrial genomes [\[27, 36\]](#). However, according to Ju et al. [\[27\]](#), the same mutational signature of the mitochondrial genome has been operative in the germline over evolutionary time and has therefore already acted on potentially synonymous sites, meaning that any new somatic changes are much less likely to be silent. Ju et al. concluded that despite the apparent high ratio of missense to silent mutations, the vast majority of mtDNA mutations were passengers with no convincing evidence for positive selection of missense mutations [\[27\]](#).

Half of the patients shared mutations between benign and cancerous tissue samples, yielding further support for the hypothesis that many putative somatic mutations are in fact low-

level heteroplasmies undetected in the germline that have undergone clonal expansion in the tumor [28]. In addition, for three patients we analyzed distinct areas within the same tumor entity, because these areas differed from each other from a histopathologically perspective. Within each patient, the different tumor areas shared mutations with each other, but also exhibited mutations that were not found in other tumor areas. This discovery yielded further support for the clonal expansion hypothesis [28], but also underscored the impression that the mutations are acquired by the tumor itself [37]. Our observations on mtDNA mutations in oral cancer were similar to those made by a study on nuclear DNA mutations in paired benign and cancerous OSCC tissue samples: Izumchenko and colleagues reported that *NOTCH1* mutations were found in 54% of primary OSCC and 60% of pre-malignant lesions. Additionally, almost 60% of leukoplakia patients with mutated *NOTCH1* carried mutations that were also identified in OSCC, indicating an important role of these clonal events in the progression of early neoplasms [38].

Four out of six patients who developed a tumor recurrence showed mutations in the recurrence that had also been observed in the index tumor. This can on the one hand be explained by clonal expansion of low-level germline heteroplasmies as described before [28]. On the other hand also the "field cancerization" concept [39] could explain the phenomenon of recurring mutations in tumor relapse tissues [40]. This concept describes a pathogenic pathway for the development of multiple OSCCs arising in different areas of the oral cavity on the basis that OSCCs are often surrounded by genetically altered cells [41, 42]. In addition to our study, another recent study based on mtDNA sequencing of OSCC samples supported this hypothesis [43].

Finally, patients with tumor metastases in their lymph nodes exhibited—on the one hand—the same somatic mtDNA mutations in the primary tumor and in their lymph node metastases and—on the other hand—showed to a great extent an increase of mutant heteroplasmy from the primary tumor to the metastases, suggesting either a selection towards the mutated mitochondrial genome during tumor progression or a higher density of tumor cells in the lymph nodes.

Another interesting observation was that tumor samples showed a higher tendency to exhibit length heteroplasmy than their corresponding benign tissue samples. For non-coding regions such as the mitochondrial control region, these additional insertions in homopolymeric tracts (>4 bp) can be explained by a lack of accuracy of the polymerase gamma leading to slippage during replication and reflect the common mtDNA mutational pattern in tumor cells [44–47]. On the contrary, only one frame-shift causing length heteroplasmy in MT-ND5 was observed in two tumor samples. This is in line with the observation of greater constraints on mitochondrial genomes carrying protein-inactivating mutations. Cancer cells carrying such deleterious mutations are at a selective disadvantage and hence do not contribute to clonal expansions, underlining the importance of functional mitochondria to cancer cells [27].

In our previous study on prostate cancer, we found that the presence of somatic mutations in transfer RNAs (tRNAs) was associated with elevated PSA levels [14], and interestingly, mtDNA mutations in tRNA regions were also described to be correlated with tumor differentiation in OSCC [16]. In addition, the observation of an increase of the degree of heteroplasmy in a tumor that originated in the prostate gland and metastasized in the seminal vesicle [14] is in line with the present-study observation of an increase of mtDNA heteroplasmy from primary tumor to lymph node metastases. Moreover, the reappearance of heteroplasmies from primary oral carcinoma in recurrences has been described in mtDNA control region sequences [43], thus hinting to a common clonal origin of malignant cells. Taken together, our recent study improves the picture of previous mtDNA studies on OSCC [16, 43, 48–50] and breast cancer [36] and sheds more light on the complex mechanism of tumor development and differentiation.

In summary, our study has several strengths: (1) it targets the entire mitochondrial genome in various types of tissue from the same oral cancer patient over a median follow-up time of 30.5 month with both Sanger and NGS procedures; (2) the high coverage of sequence reads per base (~35,000) obtained with NGS allowed a detection of mtDNA heteroplasmy at the 1% level; and (3) the longitudinal design of our study targeting various types of tissue of the same oral cancer patient at several time points increases the power of our study. The small sample size (28 OSCC patients) and the fact that blood samples were not collected from all patients however could be considered as limitations of our study.

Materials and Methods

Ethics statement

This study was approved by the ethics committee of the Medical University of Innsbruck (study code UN3564).

Patient data

Tissue samples were prospectively collected starting in June 2009. Patients aged 18 years or older, who had signed an informed consent, were consecutively included in the study. Patients with an oral squamous cell carcinoma that had been treated with chemotherapy or radiotherapy before surgery were excluded. The following clinical parameters were collected: age; sex; height; weight; smoking behavior; development of a tumor relapse during follow-up; development of metastasis; and TNM-stadium. Patients were followed for on median 30.5 months after tumor diagnosis (range 1–54 months).

Tissue samples and DNA extraction

A pathologist determined benign and malign regions on HE-stained slides from fresh frozen tissue. Macrodissection was performed on consecutive slides (10 μ m slides with an average area of 10–15 cm² of cancer and of benign tissue, respectively, which corresponded to ~5 μ g of DNA each). Then DNA was extracted from 85 tissue samples belonging to 28 patients with an EZ1 advanced Workstation with the EZ1 DNA tissue kit (QIAGEN, Hilden, Germany) and quantified with on a Tecan Infinity M200 with a Nano Quant Plate (Tecan Group Ltd. Männedorf, Switzerland). Tissue samples were taken from the primary tumors, the resection margins, the recurrences, the resection edges of the recurrences, blood, tumor areas in lymph nodes, benign areas in lymph nodes, from one second primary tumor and from one dysplasia ([S1 Table](#)).

Sanger sequencing

All samples were Sanger sequenced using the high-quality protocol described in Kloss-Brandstätter et al. 2010 [14] with an average coverage of three to five sequence electropherograms per base pair. Electropherograms were aligned to the revised Cambridge Reference Sequence (rCRS; NC_012920) [24] with Sequencher (v5, GeneCodes, Ann Arbor, MI). Similar to a study on nuclear DNA mutations [23], the limit of detection for Sanger sequencing was identified at 10% of mutant mtDNA to the reference genome rCRS [24]. There, a nucleotide position was considered heteroplasmic, if a secondary peak of more than about 10% of the height of the primary peak was present. The exact proportion of a heteroplasmic mixture was extracted from each electropherograms covering the position of interest by a software component of Sequencher. Then, the mean value of the estimated proportions was calculated and this value was used for further analyses. Every mtDNA genome was evaluated independently by two

well-experienced mtDNA technicians and validated by a senior mtDNA scientist with the mtDNA management software eCOMPAGT [51].

Next-Generation Sequencing (NGS)

All samples were subjected to NGS at AROS Applied Biotechnology (Aarhus, Denmark) on an Illumina HiSeq 2500 with an average coverage of 35,000 reads per base pair. The mitochondrial genome was initially enriched by long-range PCR-amplification of two overlapping amplicons [14] and then quantified using the Quant-iT dsDNA Broad-Range Assay Kit on the Qubit Fluorometer (Life Technologies Corporation).

Indexed paired-end DNA libraries were prepared with the TruSeq DNA HT Sample Prep Kit (Illumina, Inc.). Therefore, amplicons were fragmented to 300 bp using Covaris Adaptive Focused Acoustics technology (COVARIS, Inc.) and purified with AMPure XP beads. After end repair, 3'-adenylation, and adapter ligation, DNA samples were enriched by PCR following the TruSeq DNA HT Sample Prep Kit protocol. The libraries were then quantified using KAPA Library Quantification Kits (Kapa Biosystems). All indexed DNA libraries were pooled together with equal molar ratios and were sequenced in a single lane of one flow cell on an Illumina HiSeq 2500 using 100-bp paired-end read chemistry.

The paired-end fastq files from the Illumina HiSeq 2500 were analysed with our highly parallelized in-house-pipeline based on Cloudgene ([52] and manuscript in preparation) called mtDNA-Server (<http://mtdna-server.uibk.ac.at>). In a first step the data quality was controlled by creating reports in order to verify values such as the “per base sequence quality”, “per base N content” or the “sequence length distribution”. In a second step the reads were aligned with BWA [53] to the rCRS. From the resulting BAM-files [54], the bases for each position relative to the rCRS were extracted, whereby only bases with a PHRED-score ≥ 30 and a mapping quality ≥ 30 were used for the heteroplasmy detection. After counting the frequencies of each nucleotide per position, only such ones were marked as heteroplasmies which exceed a certain threshold, which can be defined by the user. Having a mean coverage of over 35,000 throughout our samples after applying the filters, we set the detection limit for point heteroplasmies to 1%, meaning that a base was called heteroplasmic, if the minor component amounted to at least 1% for both forward and reverse reads separately and confirmed each other.

Quality management and haplogroup determination

The mtDNA haplotypes were affiliated to haplogroups with HaploGrep [29] following Phylo-tree Build 16 [55, 56]. For quality assurance and in order to exclude contamination as potential source for the observed heteroplasmies, entire mtDNA profiles were generated from the oral surgeon, the pathologist and five lab technicians (Lab005 –Lab011) in addition to those four technicians (Lab001 –Lab004) genotyped for the prostate cancer study [14]. To do so, two profiles per sample were generated, each containing the shared variants (indicating no contamination happened), as well as the major and minor components of a heteroplasmic site respectively, as suggested by Avital et al. [57]. This way we could exclude or resequence samples, depending on the amount of remaining extracted mtDNA.

Determination of the lower detection limit of heteroplasmies

In order to find the detection threshold for heteroplasmy, four samples with pre-defined mixtures of the DNA of two lab technicians were subjected to both Sanger and NGS. After exact quantification of the entire DNA content (including nuclear DNA), the samples were mixed according to the following ratios: 1+1; 1+9; 1+49 and 1+99. Then, the levels of heteroplasmy

on the sites where the two lab members differed from each other where assessed with Sanger and next-generation sequencing ([S4 Table](#)).

Statistical and bioinformatic methods

Pearson's chi-square test was applied to 2x2 contingency tables. The Pearson product-moment correlation coefficient ("Pearson's r ") was calculated as a measure of the correlation between two continuous variables.

In order to compare Sanger and next-generation sequencing technologies for the quantification of point heteroplasmy, the mixture ratios from Sanger electropherograms were exported with Sequencher and extracted with our in-house software from NGS reads. For all heteroplasmic positions, a Bland-Altman plot was generated for the comparison of the two different heteroplasmy measurements [26]. 139 heteroplasmic mutations were below the detection threshold of Sanger sequencing, and therefore their heteroplasmy values were set to the lower limit of detection divided by the square root of two, i.e. $10/1.414 = 7.07$.

Each of the samples was represented on the graph by assigning the mean of the two measurements as the abscissa value, and the difference between the two values as the ordinate value. The mean difference reflected the estimated bias and was indicated as green line in the Bland-Altman plot. The standard deviation (SD) of the differences measured the random fluctuations around this mean. Therefore the 95% limits of agreement (average difference $\pm 1.96SD$ of the difference) were computed and plotted as red, dotted lines.

In order to identify a systematic difference between the measurements (i.e., a fixed bias), a one-sample t-test was applied to infer whether the mean value of the difference differed significantly from 0.

For comparing the frequencies of heteroplasmies at a certain level between benign and tumor samples, the numbers of heteroplasmies at a certain mixture level (e.g. 6.0–6.9%, 7.0–7.9%, etc.) were counted from all benign samples and from all tumor samples. We did not differentiate between heteroplasmies that were shared between benign and cancerous tissue samples and heteroplasmies, which occurred only in one kind of tissue. Instead, all heteroplasmic mutations were counted. Then, the frequency distributions from tumor samples and from benign samples were tested for equality with a chi-square test.

In order to evaluate the association between the occurrences of heteroplasmies in relation to gene size, the relative gene sizes were calculated by dividing the total number of base pairs per gene by the number of base pairs of the entire mitochondrial genome. Then, the number of heteroplasmies per gene was divided by the total number of heteroplasmies. Finally, these two numbers were tested for independency with a Pearson correlation analysis.

For the comparison of phylogenetic weights of heteroplasmic sites as estimated with HaploGrep [29] or EMMA [30] between different types of tissue samples, t-tests were applied. Statistical analyses were performed with IBM SPSS Statistics (version 22).

The immediate consequences (i.e. if the variants were synonymous or caused amino acid exchanges) of the observed mutations were assessed with MitoMaster [58] and MitImpact [59]. The conservation index (C.I.) was obtained from MitoTool [60]. The highest C.I. is 1, which means this site is completely conservative in 43 primate species. A C.I.-value of 0.651 indicates that 65.1% of 43 primate species have the same allele with the queried variant. MutPred [61] and PolyPhen-2 [62] were applied for predicting the functional consequences of non-synonymous mtDNA sequence variants. The number of reported NUMTs on each site was inferred from Li and colleagues [63]. In addition, the following bioinformatic tools were applied and the results collected via MitImpact [59]: FATHMM for the prediction of the functional effects of protein missense variants [64], PROVEAN to predict the functional effect of single or

multiple amino acid substitutions, insertions and deletion [65], CAROL [66] Condel [67], PhyloP [68], and PhastCons [69].

Supporting Information

S1 Fig. Mutational distribution across the mitochondrial genome in oral cancer patients.
(DOCX)

S1 Table. Description of samples and analyzed tissue.
(XLSX)

S2 Table. Point and length heteroplasmic mtDNA mutations in different tissues from 28 oral cancer patients including bioinformatic predictions on pathogenicity and functionality.
(XLSX)

S3 Table. All heteroplasmic point mutations found with NGS before filtering; any interested reader is welcome to contact us for receiving the original raw data as FASTQ-files.
(XLSX)

S4 Table. Four different mixtures of mtDNA for validation purposes.
(XLSX)

S5 Table. Point heteroplasmic mtDNA mutations in 28 benign oral tissues.
(XLSX)

S6 Table. Point heteroplasmic mtDNA mutations in 28 oral cancer tissues.
(XLSX)

S7 Table. Comparison of two different NGS runs for sample MKG04_benign.
(XLSX)

Acknowledgments

The project was supported by grants of both the Austrian Cancer Society/Tirol and the Tiroler Wissenschaftsfond, which were awarded to A. Kloss-Brandstätter. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Claudia Lamina (Division of Genetic Epidemiology, Innsbruck Medical University) for statistical and scientific advice.

Author Contributions

Conceived and designed the experiments: AKB HK CWH MR FK FRK GS. Performed the experiments: AKB HW GE LF SS DP CS AS IS FRK. Analyzed the data: AKB HW GE LF SS DP. Contributed reagents/materials/analysis tools: AKB HW GE LF SS HK CWK MR FK FRK. Wrote the paper: AKB HW GE LF FK FRK. Operated the patients: FRK MR.

References

1. Perez-Sayans M, Somoza-Martin JM, Barros-Angueira F, Reboiras-Lopez MD, Gandara Rey JM, Garcia-Garcia A. Genetic and molecular alterations associated with oral squamous cell cancer (Review). *Oncol Rep.* 2009; 22(6):1277–82. PMID: [19885577](#)
2. Heck JE, Berthiller J, Vaccarella S, Winn DM, Smith EM, Shan'gina O, et al. Sexual behaviours and the risk of head and neck cancers: a pooled analysis in the International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Int J Epidemiol.* 2010; 39(1):166–81. dyp350 [pii]; doi: [10.1093/ije/dyp350](#) PMID: [20022926](#)

3. Miller CS, Henry RG, Rayens MK. Disparities in risk of and survival from oropharyngeal squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2003; 95(5):570–5. doi: [10.1067/moe.2003.108](https://doi.org/10.1067/moe.2003.108) S1079210402917087 [pii]. PMID: [12738948](https://pubmed.ncbi.nlm.nih.gov/12738948/)
4. Wolff D, Hassfeld S, Hofele C. Influence of marginal and segmental mandibular resection on the survival rate in patients with squamous cell carcinoma of the inferior parts of the oral cavity. *J Craniomaxillofac Surg.* 2004; 32(5):318–23. doi: [10.1016/j.jcms.2004.05.005](https://doi.org/10.1016/j.jcms.2004.05.005) S1010-5182(04)00086-1 [pii]. PMID: [15458675](https://pubmed.ncbi.nlm.nih.gov/15458675/)
5. Shutt TE, Shadel GS. A compendium of human mitochondrial gene expression machinery with links to disease. *Environ Mol Mutagen.* 2010; 51(5):360–79. doi: [10.1002/em.20571](https://doi.org/10.1002/em.20571) PMID: [20544879](https://pubmed.ncbi.nlm.nih.gov/20544879/)
6. Riemer J, Bulleid N, Herrmann JM. Disulfide formation in the ER and mitochondria: two solutions to a common process. *Science.* 2009; 324(5932):1284–7. 324/5932/1284 [pii]; doi: [10.1126/science.1170653](https://doi.org/10.1126/science.1170653) PMID: [19498160](https://pubmed.ncbi.nlm.nih.gov/19498160/)
7. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet.* 2009; 84(6):740–59. S0002-9297(09)00163-3 [pii]; doi: [10.1016/j.ajhg.2009.05.001](https://doi.org/10.1016/j.ajhg.2009.05.001) PMID: [19500773](https://pubmed.ncbi.nlm.nih.gov/19500773/)
8. Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, 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.* 2009; 68(5):516–27. doi: [10.1007/s00239-009-9227-4](https://doi.org/10.1007/s00239-009-9227-4) PMID: [19407924](https://pubmed.ncbi.nlm.nih.gov/19407924/)
9. Samuels DC, Li C, Li B, Song Z, Torstenson E, Boyd CH, et al. Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. *PLoS Genet.* 2013; 9(11):e1003929. doi: [10.1371/journal.pgen.1003929](https://doi.org/10.1371/journal.pgen.1003929) PGENETICS-D-13-01891 [pii]. PMID: [24244193](https://pubmed.ncbi.nlm.nih.gov/24244193/)
10. Ramos A, Santos C, Mateiu L, Gonzalez MM, Alvarez L, Azevedo L, et al. Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. *PLoS ONE.* 2013; 8(10):e74636. doi: [10.1371/journal.pone.0074636](https://doi.org/10.1371/journal.pone.0074636) PONE-D-13-11473 [pii]. PMID: [24098342](https://pubmed.ncbi.nlm.nih.gov/24098342/)
11. Freyer C, Cree LM, Mourier A, Stewart JB, Koolmeister C, Milenkovic D, et al. Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. *Nat Genet.* 2012; 44(11):1282–5. ng.2427 [pii]; doi: [10.1038/ng.2427](https://doi.org/10.1038/ng.2427) PMID: [23042113](https://pubmed.ncbi.nlm.nih.gov/23042113/)
12. Fendt L, Niederstätter H, Huber G, Zelger B, Dunser M, Seifarth C, et al. Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection. *Breast Cancer Res Treat.* 2011; 128(2):327–36. doi: [10.1007/s10549-010-1092-8](https://doi.org/10.1007/s10549-010-1092-8) PMID: [20697806](https://pubmed.ncbi.nlm.nih.gov/20697806/)
13. Santos C, Martinez M, Lima M, Hao YJ, Simoes N, Montiel R. Mitochondrial DNA mutations in cancer: a review. *Curr Top Med Chem.* 2008; 8(15):1351–66. PMID: [18991723](https://pubmed.ncbi.nlm.nih.gov/18991723/)
14. Kloss-Brandstätter A, Schäfer G, Erhart G, Hüttenhofer A, Coassin S, Seifarth C, et al. Somatic Mutations throughout the Entire Mitochondrial Genome Are Associated with Elevated PSA Levels in Prostate Cancer Patients. *Am J Hum Genet.* 2010; 87(6):802–12. doi: [10.1016/j.ajhg.2010.11.001](https://doi.org/10.1016/j.ajhg.2010.11.001) PMID: [21129724](https://pubmed.ncbi.nlm.nih.gov/21129724/)
15. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature.* 2010; 464(7288):610–4. nature08802 [pii]; doi: [10.1038/nature08802](https://doi.org/10.1038/nature08802) PMID: [20200521](https://pubmed.ncbi.nlm.nih.gov/20200521/)
16. Lai CH, Huang SF, Liao CT, Chen IH, Wang HM, Hsieh LL. Clinical significance in oral cavity squamous cell carcinoma of pathogenic somatic mitochondrial mutations. *PLoS ONE.* 2013; 8(6):e65578. doi: [10.1371/journal.pone.0065578](https://doi.org/10.1371/journal.pone.0065578) PONE-D-13-05358 [pii]. PMID: [23799027](https://pubmed.ncbi.nlm.nih.gov/23799027/)
17. Carr IM, Robinson JI, Dimitriou R, Markham AF, Morgan AW, Bonthron DT. Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics.* 2009; 25(24):3244–50. btp583 [pii]; doi: [10.1093/bioinformatics/btp583](https://doi.org/10.1093/bioinformatics/btp583) PMID: [19819885](https://pubmed.ncbi.nlm.nih.gov/19819885/)
18. Tang S, Wang J, Zhang VW, Li FY, Landsverk M, Cui H, et al. Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects. *Hum Mutat.* 2013; 34(6):882–93. doi: [10.1002/humu.22307](https://doi.org/10.1002/humu.22307) PMID: [23463613](https://pubmed.ncbi.nlm.nih.gov/23463613/)
19. Krjtskov K, Koltsina M, Grand K, Vosa U, Sauk M, Tonisson N, et al. Tissue-specific mitochondrial heteroplasmy at position 16,093 within the same individual. *Curr Genet.* 2014; 60(1):11–6. doi: [10.1007/s00294-013-0398-6](https://doi.org/10.1007/s00294-013-0398-6) PMID: [23842853](https://pubmed.ncbi.nlm.nih.gov/23842853/)
20. Li M, Schönberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet.* 2010; 87(2):237–49. S0002-9297(10)00370-8 [pii]; doi: [10.1016/j.ajhg.2010.07.014](https://doi.org/10.1016/j.ajhg.2010.07.014) PMID: [20696290](https://pubmed.ncbi.nlm.nih.gov/20696290/)
21. Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet.* 2011; 19(9):988–94. ejhg201162 [pii]; doi: [10.1038/ejhg.2011.62](https://doi.org/10.1038/ejhg.2011.62) PMID: [21487439](https://pubmed.ncbi.nlm.nih.gov/21487439/)

22. Zhang W, Cui H, Wong LJ. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin Chem*. 2012; 58(9):1322–31. [clinchem.2011.181438](https://doi.org/10.1373/clinchem.2011.181438) [pii]; doi: [10.1373/clinchem.2011.181438](https://doi.org/10.1373/clinchem.2011.181438) PMID: [22777720](https://pubmed.ncbi.nlm.nih.gov/22777720/)
23. Aydin S, Dekairelle AF, Ambroise J, Durant JF, Heusterspreute M, Guiot Y, et al. Unambiguous detection of multiple TP53 gene mutations in AAN-associated urothelial cancer in Belgium using laser capture microdissection. *PLoS ONE*. 2014; 9(9):e106301. doi: [10.1371/journal.pone.0106301](https://doi.org/10.1371/journal.pone.0106301) PONE-D-14-24524 [pii]. PMID: [25184754](https://pubmed.ncbi.nlm.nih.gov/25184754/)
24. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999; 23(2):147–. PMID: [10508508](https://pubmed.ncbi.nlm.nih.gov/10508508/)
25. Croghan CW, Egeghy PP. Methods of Dealing with Values Below the Limit of Detection using SAS 2003 [updated 2003]. Available: <http://analytics.ncsu.edu/sesug/2003/SD08-Croghan.pdf>.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1(8476):307–10. S0140-6736(86)90837-8 [pii]. PMID: [2868172](https://pubmed.ncbi.nlm.nih.gov/2868172/)
27. Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife*. 2014; 3(e029535). doi: [10.7554/eLife.02935](https://doi.org/10.7554/eLife.02935)
28. Payne BA, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, et al. Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet*. 2013; 22(2):384–90. [dds435](https://doi.org/10.1093/hmg/dds435) [pii]; doi: [10.1093/hmg/dds435](https://doi.org/10.1093/hmg/dds435) PMID: [23077218](https://pubmed.ncbi.nlm.nih.gov/23077218/)
29. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, et al. HaploGrep: A Fast and Reliable Algorithm for Automatic Classification of Mitochondrial DNA Haplogroups. *Hum Mutat*. 2011; 32(1):25–32. doi: [10.1002/humu.21382](https://doi.org/10.1002/humu.21382) PMID: [20960467](https://pubmed.ncbi.nlm.nih.gov/20960467/)
30. Röck AW, Dür A, vO M., Parson W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Sci Int Genet*. 2013; 7(6):601–9. S1872-4973(13)00155-5 [pii]; doi: [10.1016/j.fsigen.2013.07.005](https://doi.org/10.1016/j.fsigen.2013.07.005) PMID: [23948335](https://pubmed.ncbi.nlm.nih.gov/23948335/)
31. Brandstätter A, Sängler T, Lutz-Bonengel S, Parson W, Béraud-Colomb E, Wen B, et al. Phantom mutation hotspots in human mitochondrial DNA. *Electrophoresis*. 2005; 26(18):3414–29. PMID: [16167362](https://pubmed.ncbi.nlm.nih.gov/16167362/)
32. Li H. Improving SNP discovery by base alignment quality. *Bioinformatics*. 2011; 27(8):1157–8. [btr076](https://doi.org/10.1093/bioinformatics/btr076) [pii]; doi: [10.1093/bioinformatics/btr076](https://doi.org/10.1093/bioinformatics/btr076) PMID: [21320865](https://pubmed.ncbi.nlm.nih.gov/21320865/)
33. Naue J, Horer S, Sanger T, Strobl C, Hatzer-Grubwieser P, Parson W, et al. Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion*. 2015; 20:82–94. doi: [10.1016/j.mito.2014.12.002](https://doi.org/10.1016/j.mito.2014.12.002) PMID: [25526677](https://pubmed.ncbi.nlm.nih.gov/25526677/).
34. Li M, Schroder R, Ni S, Madea B, Stoneking M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc Natl Acad Sci U S A*. 2015; 112(8):2491–6. [1419651112](https://doi.org/10.1073/pnas.1419651112) [pii]; doi: [10.1073/pnas.1419651112](https://doi.org/10.1073/pnas.1419651112) PMID: [25675502](https://pubmed.ncbi.nlm.nih.gov/25675502/)
35. Gardner K, Payne BA, Horvath R, Chinnery PF. Use of stereotypical mutational motifs to define resolution limits for the ultra-deep resequencing of mitochondrial DNA. *Eur J Hum Genet*. 2015; 23(3):413–5. [ejhg201496](https://doi.org/10.1038/ejhg.2014.96) [pii]; doi: [10.1038/ejhg.2014.96](https://doi.org/10.1038/ejhg.2014.96) PMID: [24896153](https://pubmed.ncbi.nlm.nih.gov/24896153/)
36. McMahon S, Laframboise T. Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates. *Carcinogenesis*. 2014; 35(5):1046–54. [bgu012](https://doi.org/10.1093/carcin/bgu012) [pii]; doi: [10.1093/carcin/bgu012](https://doi.org/10.1093/carcin/bgu012) PMID: [24442641](https://pubmed.ncbi.nlm.nih.gov/24442641/)
37. Wallace DC, Chalkia D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol*. 2013; 5(11):a021220. [5/11/a021220](https://doi.org/10.1101/cshperspect.a021220) [pii]; doi: [10.1101/cshperspect.a021220](https://doi.org/10.1101/cshperspect.a021220) PMID: [24186072](https://pubmed.ncbi.nlm.nih.gov/24186072/)
38. Izumchenko E, Sun K, Jones S, Brait M, Agrawal N, Koch WM, et al. Notch1 mutations are drivers of oral tumorigenesis. *Cancer Prev Res (Phila)*. 2014. [1940–6207.CAPR-14-0257](https://doi.org/10.1158/1940-6207.CAPR-14-0257) [pii]; doi: [10.1158/1940-6207.CAPR-14-0257](https://doi.org/10.1158/1940-6207.CAPR-14-0257)
39. SLAUGHTER DP, SOUTHWICK HW, SMEJKAL W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer*. 1953; 6(5):963–8. PMID: [13094644](https://pubmed.ncbi.nlm.nih.gov/13094644/)
40. Foschini MP, Morandi L, Marchetti C, Cocchi R, Eusebi LH, Farnedi A, et al. Cancerization of cutaneous flap reconstruction for oral squamous cell carcinoma: report of three cases studied with the mtDNA D-loop sequence analysis. *Histopathology*. 2011; 58(3):361–7. doi: [10.1111/j.1365-2559.2011.03754.x](https://doi.org/10.1111/j.1365-2559.2011.03754.x) PMID: [21255069](https://pubmed.ncbi.nlm.nih.gov/21255069/)
41. Bedi GC, Westra WH, Gabrielson E, Koch W, Sidransky D. Multiple head and neck tumors: evidence for a common clonal origin. *Cancer Res*. 1996; 56(11):2484–7. PMID: [8653681](https://pubmed.ncbi.nlm.nih.gov/8653681/)
42. Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res*. 2003; 63(8):1727–30. PMID: [12702551](https://pubmed.ncbi.nlm.nih.gov/12702551/)

43. Montebugnoli L, Leonardi E, Morandi L, Farnedi A, Gissi DB, Marchetti C, et al. Genetic relationship between multiple squamous cell carcinomas arising in the oral cavity. *Head Neck*. 2014; 36(1):94–100. doi: [10.1002/hed.23259](https://doi.org/10.1002/hed.23259) PMID: [23633084](https://pubmed.ncbi.nlm.nih.gov/23633084/)
44. Zheng W, Khrapko K, Collier HA, Thilly WG, Copeland WC. Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors. *Mutat Res*. 2006; 599(1–2):11–20. S0027-5107(06)00027-3 [pii]; doi: [10.1016/j.mrfmmm.2005.12.012](https://doi.org/10.1016/j.mrfmmm.2005.12.012) PMID: [16490220](https://pubmed.ncbi.nlm.nih.gov/16490220/)
45. Kunkel TA, Soni A. Exonucleolytic proofreading enhances the fidelity of DNA synthesis by chick embryo DNA polymerase-gamma. *J Biol Chem*. 1988; 263(9):4450–9. PMID: [2831231](https://pubmed.ncbi.nlm.nih.gov/2831231/)
46. Bailey CM, Anderson KS. A mechanistic view of human mitochondrial DNA polymerase gamma: providing insight into drug toxicity and mitochondrial disease. *Biochim Biophys Acta*. 2010; 1804(5):1213–22. S1570-9639(10)00009-9 [pii]; doi: [10.1016/j.bbapap.2010.01.007](https://doi.org/10.1016/j.bbapap.2010.01.007) PMID: [20083238](https://pubmed.ncbi.nlm.nih.gov/20083238/)
47. Hudson G, Chinnery PF. Mitochondrial DNA polymerase-gamma and human disease. *Hum Mol Genet*. 2006; 15 Spec No 2:R244–R52. 15/suppl_2/R244 [pii]; doi: [10.1093/hmg/ddl233](https://doi.org/10.1093/hmg/ddl233) PMID: [16987890](https://pubmed.ncbi.nlm.nih.gov/16987890/)
48. Mondal R, Ghosh SK, Choudhury JH, Seram A, Sinha K, Hussain M, et al. Mitochondrial DNA copy number and risk of oral cancer: a report from Northeast India. *PLoS ONE*. 2013; 8(3):e57771. doi: [10.1371/journal.pone.0057771](https://doi.org/10.1371/journal.pone.0057771) PONE-D-12-37543 [pii]. PMID: [23469236](https://pubmed.ncbi.nlm.nih.gov/23469236/)
49. Mondal R, Ghosh SK. Accumulation of mutations over the complete mitochondrial genome in tobacco-related oral cancer from northeast India. *Mitochondrial DNA*. 2013; 24(4):432–9. doi: [10.3109/19401736.2012.760551](https://doi.org/10.3109/19401736.2012.760551) PMID: [23350716](https://pubmed.ncbi.nlm.nih.gov/23350716/)
50. Challen C, Brown H, Cai C, Betts G, Paterson I, Sloan P, et al. Mitochondrial DNA mutations in head and neck cancer are infrequent and lack prognostic utility. *Br J Cancer*. 2011; 104(8):1319–24. bjc201196 [pii]; doi: [10.1038/bjc.2011.96](https://doi.org/10.1038/bjc.2011.96) PMID: [21427725](https://pubmed.ncbi.nlm.nih.gov/21427725/)
51. Weissensteiner H, Schönherr S, Specht G, Kronenberg F, Brandstätter A. eCOMPAGT integrates mtDNA: import, validation and export of mitochondrial DNA profiles for population genetics, tumour dynamics and genotype-phenotype association studies. *BMC Bioinformatics*. 2010; 11:122. doi: [10.1186/1471-2105-11-122](https://doi.org/10.1186/1471-2105-11-122) PMID: [20214782](https://pubmed.ncbi.nlm.nih.gov/20214782/)
52. Schönherr S, Forer L, Weissensteiner H, Kronenberg F, Specht G, Kloss-Brandstätter A. Cloudbase: a graphical execution platform for MapReduce programs on private and public clouds. *BMC Bioinformatics*. 2012; 13:200. 1471-2105-13-200 [pii]; doi: [10.1186/1471-2105-13-200](https://doi.org/10.1186/1471-2105-13-200) PMID: [22888776](https://pubmed.ncbi.nlm.nih.gov/22888776/)
53. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010; 26(5):589–95. btp698 [pii]; doi: [10.1093/bioinformatics/btp698](https://doi.org/10.1093/bioinformatics/btp698) PMID: [20080505](https://pubmed.ncbi.nlm.nih.gov/20080505/)
54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. btp352 [pii]; doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
55. van Oven M. Revision of the mtDNA tree and corresponding haplogroup nomenclature. *Proc Natl Acad Sci U S A*. 2010; 107(11):E38–E9. 107/11/E38 [pii]; doi: [10.1073/pnas.0915120107](https://doi.org/10.1073/pnas.0915120107) PMID: [20304802](https://pubmed.ncbi.nlm.nih.gov/20304802/)
56. van Oven M, Kayser M. Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Human Mutation*. 2009; 30(2):E386–E94. doi: [10.1002/humu.20921](https://doi.org/10.1002/humu.20921) PMID: [18853457](https://pubmed.ncbi.nlm.nih.gov/18853457/)
57. Avital G, Buchshtav M, Zhidkov I, Tuval FJ, Dadon S, Rubin E, et al. Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. *Hum Mol Genet*. 2012; 21(19):4214–24. dds245 [pii]; doi: [10.1093/hmg/dds245](https://doi.org/10.1093/hmg/dds245) PMID: [22736028](https://pubmed.ncbi.nlm.nih.gov/22736028/)
58. Brandon MC, Ruiz-Pesini E, Mishmar D, Procaccio V, Lott MT, Nguyen KC, et al. MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum Mutat*. 2009; 30(1):1–6. doi: [10.1002/humu.20801](https://doi.org/10.1002/humu.20801) PMID: [18566966](https://pubmed.ncbi.nlm.nih.gov/18566966/)
59. Castellana S, Ronai J, Mazza T. MitlImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Hum Mutat*. 2015; 36(2):E2413–22. doi: [10.1002/humu.22720](https://doi.org/10.1002/humu.22720) PMID: [25516408](https://pubmed.ncbi.nlm.nih.gov/25516408/)
60. Fan L, Yao YG. MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion*. 2011; 11(2):351–6. S1567-7249(10)00168-6 [pii]; doi: [10.1016/j.mito.2010.09.013](https://doi.org/10.1016/j.mito.2010.09.013) PMID: [20933105](https://pubmed.ncbi.nlm.nih.gov/20933105/)
61. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009; 25(21):2744–50. btp528 [pii]; doi: [10.1093/bioinformatics/btp528](https://doi.org/10.1093/bioinformatics/btp528) PMID: [19734154](https://pubmed.ncbi.nlm.nih.gov/19734154/)
62. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7(4):248–9. nmeth0410-248 [pii]; doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248) PMID: [20354512](https://pubmed.ncbi.nlm.nih.gov/20354512/)
63. Li M, Schroeder R, Ko A, Stoneking M. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res*. 2012; 40(18):e137. doi: [10.1093/nar/gks499](https://doi.org/10.1093/nar/gks499) PMID: [22649055](https://pubmed.ncbi.nlm.nih.gov/22649055/); PubMed Central PMCID: [PMC3467033](https://pubmed.ncbi.nlm.nih.gov/PMC3467033/)

64. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat.* 2013; 34(1):57–65. doi: [10.1002/humu.22225](https://doi.org/10.1002/humu.22225) PMID: [23033316](https://pubmed.ncbi.nlm.nih.gov/23033316/); PubMed Central PMCID: PMC3558800.
65. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics.* 2015. doi: [10.1093/bioinformatics/btv195](https://doi.org/10.1093/bioinformatics/btv195) PMID: [25851949](https://pubmed.ncbi.nlm.nih.gov/25851949/).
66. Lopes MC, Joyce C, Ritchie GR, John SL, Cunningham F, Asimit J, et al. A combined functional annotation score for non-synonymous variants. *Hum Hered.* 2012; 73(1):47–51. doi: [10.1159/000334984](https://doi.org/10.1159/000334984) PMID: [22261837](https://pubmed.ncbi.nlm.nih.gov/22261837/); PubMed Central PMCID: PMC3390741.
67. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet.* 2011; 88(4):440–9. doi: [10.1016/j.ajhg.2011.03.004](https://doi.org/10.1016/j.ajhg.2011.03.004) PMID: [21457909](https://pubmed.ncbi.nlm.nih.gov/21457909/); PubMed Central PMCID: PMC3071923.
68. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 2010; 20(1):110–21. doi: [10.1101/gr.097857.109](https://doi.org/10.1101/gr.097857.109) PMID: [19858363](https://pubmed.ncbi.nlm.nih.gov/19858363/); PubMed Central PMCID: PMC2798823.
69. King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Res.* 2005; 15(8):1051–60. doi: [10.1101/gr.3642605](https://doi.org/10.1101/gr.3642605) PMID: [16024817](https://pubmed.ncbi.nlm.nih.gov/16024817/); PubMed Central PMCID: PMC1182217.