Published in final edited form as:

Methods. 2008 December ; 46(4): 269–273. doi:10.1016/j.ymeth.2008.10.005.

Single molecule PCR in mtDNA mutational analysis: genuine mutations vs. damage bypass-derived artifacts

Y Kraytsberg¹, A Nicholas¹, P Caro², and K Khrapko¹

¹ Beth Israel Deaconess Medical Center and Harvard Medical School, Boston MA

² Department of Animal Physiology-II, Complutense University, Madrid, Spain

Abstract

The area of somatic mtDNA mutation measurement is in a crisis because the methods used to quantify mtDNA mutations produce results varying by multiple orders of magnitude. The reason for these discrepancies is not clear, but given that most methods involve PCR, the prime suspect is PCR artifacts (e.g. spontaneous errors by the DNA polymerases used). In addition to simple misincorporation, another important source of artificial mutations is the conversion of chemically modified (e.g. damaged) nucleotides into mutations when bypassed by a thermostable DNA polymerase. These latter mutations are particularly difficult to account for because appropriate controls are not available. Here, we argue that single molecule PCR (smPCR) is uniquely positioned to account for these bypass-related artificial mutations and discuss the methodology involved in employing this technique.

1. Introduction

Measurements of the fractions of somatic mtDNA mutations are instrumental in evaluating the various hypotheses relating mtDNA mutations to aging and disease. Somatic mtDNA mutations may also be used as markers of cancer cells and could be used in diagnostics or recurrent cancers. Unfortunately the field of measurements of somatic mtDNA is highly controversial. Estimates of mutant fractions in the same type of tissue may vary by orders of magnitude depending on the method used. For example, estimates of the mutant fraction of young murine brain range from 0.6×10^{-6} by random mutation capture approach (1) (section 2.2) to 4×10^{-6} by direct cloning (2) (section 2.1) and to 3×10^{-4} (3) by PCR-cloning approach (section 2.3). The cause of this discrepancy may be that most of contemporary methods of mutational analysis use PCR, as PCR greatly simplifies DNA work and allows to analyse samples containing small amounts of DNA. The price for efficiency and convenience are the various PCR artifacts that introduce artificial mutations that compromise mutational analysis as long as their frequency is comparable to that of genuine mutations. Traditionally, PCR-induced mutations are considered polymerase errors. It is now being progressively recognized (4,5), that some PCR artifacts may represent damage-induced mutations, i.e. mutations resulting from the bypass of chemically damaged nucleotides by DNA polymerase. For example, a common modification, deaminated cytosine, is recognized by the Taq polymerase as uracil, which results in C to T base change (6). We will briefly review the existing methods

designed to reduce or eliminate PCR-related artefacts of mutational analysis and then will concentrate on one promising approach, the single molecule PCR.

2. Methods of somatic mtDNA mutational analysis: the different ways to cope with artifacts

With respect to handling PCR artifacts, mtDNA mutation analysis methods include (but are not limited to) the following four groups.

2.1. *Direct cloning involves isolation of mtDNA (i.e., separation from nuclear DNA, usually by isolation of mitochondria)*

DNA is then digested by restriction enzymes and cloned in bacterial vectors using standard protocols. Individual clones are then sequenced and any base pair changes in the cloned sequences are considered mutations. This approach was originally applied to mtDNA over two decades ago (7) and is occasionally used now (2,8). Obviously, direct cloning totally avoids PCR and is usually considered error-free. It still has to be proven, however, that cloning does not introduce artificial mutations: cloning involves DNA replication and possibly repair, a potentially error-prone processes. Direct cloning is quite laborious and cannot be used in cases where amount of DNA is limiting or of poor quality, which is commonplace when human tissues or single cells are analysed.

2.2. Removal of wild type templates

A group of methods strives to avoid artifacts by removing the wild type (i.e. non-mutant) molecules. This approach is based on an idea that, since genuine somatic mutants constitute only a very small proportion of molecules, most of artificial mutations (both polymerase errors and converted damage) by pure chance occur almost exclusively on wild type molecules. Removal of the latter ones prior to PCR may thus potentially improve the ratio of genuine mutants to artifacts by many orders of magnitude. So far, efficient removal of wild type was achieved either by partially denaturing electrophoresis (9) or by restriction digestion (Random mutation capture, RMC – Vermulst et al., this issue) (1). Both approaches, however, use rather short target sequences (i.e. DNA segments that are scanned for mutations), which has two drawbacks. First, physiologically important mutations and intense mutational hotspots most likely will be left outside the short target sequence which may lead to underestimates of both overall mutant fractions and physiological significance of mtDNA mutations. In agreement with this hypothesis, RMC where target sequence is particularly short, 4bp only, tends to yield the lowest estimates of mutant fractions compared to other methods. Second, with short target sequences, it is not practical to measure mutations in individual cells, because it is unlikely that mutation present in a given cell by chance falls within the target. It should be noted that artificial mutations resulting from bypassed DNA damage may affect the results of methods that remove wild type DNA as well. Indeed, some types of damage may theoretically behave as true mutations as far as wt removal is concerned, (i.e. be co-separated with mutations or fail to get digested by a restriction enzyme). This area remains poorly studied, though in some cases this possibility is being accounted for, for example, some authors specifically use artificially damaged DNA to prove that damage does not induce extra mutations (5).

2.3. PCR-cloning methods

This approach (10) includes PCR of the DNA as the first step, with subsequent cloning and sequencing of PCR products. This technique is very efficient and convenient, allows to analyse extremely small amounts of DNA, including single cells. The limitation of this approach is that any artificial mutation generated during PCR becomes indistinguishable from a genuine mutation, so the only way to account for artifacts is to use controls, such as supposedly

mutation-free DNA (e.g. cloned DNA) subject to the same procedures as the sample DNA (i.e. DNA where mutational levels are being measured). As long as mutation levels in the control are significantly lower than in the sample, the excess of mutations in the sample is assumed to be genuine mutations. Unfortunately, this type of control may not fully account for chemical damage of the sample DNA, and thus excess of mutations may represent artificial mutations resulting from bypass of damage during PCR.

2.4. Single cell mutational analysis

This method takes advantage of the phenomenon of clonal expansion of mtDNA in individual cells. Since each cell carries large numbers of mtDNA copies, mutations are generally described by their extent of intracellular heteroplasmy (the percentage of copies carrying the mutation). Interestingly, at least in some cell types, somatic mtDNA mutations manage to clonally expand from initial single mutational event to dominate mtDNA population in the cell (11),(12). Whereas mechanisms of clonal expansion are a matter of debate, from an analytical point of view, this phenomenon allows easy measurement of mtDNA mutations if single cells are analyzed. Since clonally expanded mutations are present at significant fractions, artificial mutations are not a problem because they arise at much lower frequencies and do not interfere with the analysis. This approach is very laborious and requires intact tissue rather than DNA, which is not always available. Furthermore, single cell analysis does not measure mutations that have not been expanded (or cannot be expanded). Non-expanded mutations are generally considered irrelevant to physiology of the cell because they are below physiological threshold. However, growing body of evidence implies that there may exist a special class of mutations (dominant lethal) that may kill a cell even when present at very low levels and that these mutations may play an important role in the aging process (13). The ability to measure non-expanded mutations is vital either to confirm or reject this hypothesis. Single cell mtDNA mutational analysis has been described elsewhere (14).

In conclusion, none of the above-mentioned methods is capable of measuring non-expanded mtDNA mutations without the risk of getting a poorly controlled admixture of artificial mutations.

3. Avoiding artifacts by means of smPCR: the theory

As long as DNA molecules are handled en masse, genuine somatic mutations are essentially indistinguishable from artificial base changes arising from bypass of damage during PCR. However, the behaviour of genuine mutations and artifacts becomes drastically different when one deals with single molecules, as described in detail below. This situation allows one to further identify and discriminate artifacts from genuine mutations.

The scheme illustrates the different outputs resulting from smPCR amplification of genuine double stranded mutations (B) as contrasted to polymerase errors (C) or mutations resulting from DNA damage bypass (D). Mutations resulting from damage bypass may be indistinguishable from genuine mutations if one strand of template cannot be amplified due to an impassable site (E). Panel A shows amplification of a wild type molecule. See section 3 for details.

3.1. Polymerase errors

DNA is a double stranded molecule, and any genuine mutation is present in both strands of native cellular DNA. Obviously, a mutant template DNA, when subject to PCR results in PCR product such that every molecule in the mixture carries the mutation (Fig 1B). In contrast to a genuine in vivo mutation, an artificial base substitution resulting from an error of the thermostable DNA polymerase (Figure 1C) emerges as polymerase copies one but not the other

strand of the initial template molecule. If all DNA strands in PCR reaction are copied equally efficiently, then any polymerase error that emerged at the first PCR cycle will constitute 25% of the PCR product. Obviously the 100% and the 25% of a base change can be easily discriminated by direct sequencing of the PCR product. Any polymerase error that occurs on a later PCR cycle will represent even smaller proportion of PCR products (1/8 for the second cycle, 1/16 for the third, and so on) so these later errors will be disregarded in the discussion to follow as they are even easier to discriminate.

3.2. Bypassed DNA damage converted into mutations

There are certain types of DNA damage, which may, when bypassed by DNA polymerases, result in base changes. For example, a deaminated cytosine is recognised by some polymerases, including the widely used Taq polymerase, as thymine, and an adenosine is inserted into the nascent DNA strand across from the damaged base which results in a C:G to T:A transition. Repeated misreading of a deaminated C in the original DNA template ultimately results in a 50:50 mutant:wild type mixture of PCR fragments (Figure 1D), which is still easy to distinguish from a 100% genuine mutation.

3.3. Impassable lesions – the weakness of the smPCR approach

It should be noted, however, that PCR does not necessarily start with a double-stranded DNA. Every positive PCR reaction by definition contains at least one amplifiable DNA strand, but there is no guarantee that both strands of the double stranded template can be amplified even if they are physically present in the reaction tube. DNA strands may contain nicks or otherwise “impassable” damage sites indicated by a cross (X) in Figure 1E. In practice it appears that a significant proportion of templates in a typical DNA sample cannot be amplified. This can be demonstrated by comparing of the copy number of DNA template as estimated using shorter vs. longer PCR fragments. For example, there may be 20-fold more 3kb-long than 14kb-long templates in a regular mtDNA sample (Kraytsberg et al in press). This implies that every 15-kb DNA strand probably carries a few impassable sites on average. This implies that, at least for longer templates, the most likely situation is that in a double stranded amplifiable template, only one of the strands does not contain impassable sites, and that in a smPCR reaction, there usually are several dsDNA molecules, only one of which contains one amplifiable strand and a non-amplifiable second strand, while other dsDNA molecules contain both non-amplifiable strands. This means that most probable is the scenario shown in Fig. 1E, where a misspelled modified base results in a 100% artificial mutation indistinguishable from a genuine mutation. Similarly, a polymerase error will result in a 50% artificial mutation if one of the strands of template is non-amplifiable.

3.4. How important is damage converted to mutations?

The contribution of damage bypass-related mutations is not well-studied and may critically depend on the type of the thermostable DNA polymerase used for PCR. In particular, pfu and related polymerases are highly discriminative. For example, they do not bypass certain damaged nucleotides such as deaminated cytosine (which is readily bypassed and converted into T by less-discriminative polymerases, such as Taq), thus, at least this lesion, would not result in artificial mutations in this case. It has been shown that the amount of converted damage is substantial (i.e. at least comparable to the amount of genuine mutations) when non-discriminative DNA polymerases (mitochondrial DNA polymerase gamma in this case) are used to amplify DNA (6). Less is known about the fraction of converted mutations in DNA when discriminative polymerases are used. At least some such mutations are present even when pfu and enrichment of mutated DNA achieved by partially denaturing electrophoresis are used to measure mutations as described in section 2.2. These artificial mutations were detected using asymmetric PCR, i.e. PCR initiated with several cycles of linear amplification with only one

primer followed by conventional two-primer PCR. Asymmetric PCR stage ensures that PCR product primarily originates either from one or the other of the DNA strands, depending on which of the two primers was used in the asymmetric step. The proportion of converted damage-derived mutations in mutational spectra obtained using pfu polymerase has been determined vaguely and apparently varies from sample to sample (4). It appears however that it may be still comparable to genuine mutations in typical human DNA samples. In conclusion, there is a real possibility that damage-derived artificial mutations significantly pollute observed mutational spectra and a means to estimate the fraction of these lesions is urgently needed.

3.5. Controlling damage converted to mutations in smPCR

The potential presence of damage bypass-related mutations implies that it is important to be able to perform smPCR under conditions where both strands of the initial template molecule are utilized by PCR. One way to make sure that both strands are utilized is to amplify progressively shorter PCR fragments. As the length of PCR fragment decreases, the probability that none of the strands contains an impassable lesion decreases and thus the probability that both strands of the template will participate in PCR increases and reaches a plateau where both strands are fully utilized. The fact that both strands are indeed utilized in PCR can be verified by denaturing DNA prior to aliquoting the template for single molecule PCR. If both strands are amplifiable, pre-aliquoting denaturing should double the number of molecules as measured by single molecule PCR.

4. The practice of smPCR: general considerations

This section contains a few general recommendations and tips for using the “sample protocol” presented in the next section.

4.1. Achieving single molecule level

smPCR is essentially normal PCR where template DNA is diluted to very low “limiting dilution” concentration, such that a reaction tube typically contains one amplifiable template. In a typical experiment, template is sequentially diluted and subjected to PCR in multiple wells, until a characteristic picture of single molecule PCR appears (some wells are completely empty, while others contain intense PCR products (Figure 2)). However, such a picture will appear only if PCR system used is capable of delivering high quality PCR product starting with a single template copy. If this “yes or no” situation is not achieved, and PCR rather “wanes out” in all reactions, then PCR conditions are not efficient enough for single molecule PCR and need to be adjusted as discussed below. The two major technical problems associated with the use of extremely low concentrations of template are competing adverse processes and contamination. The art of performing smPCR mainly consists of dealing with these two main problems.

Lanes containing PCR products are the reactions that happened to contain a single DNA template (black arrowhead). Note that all lanes contain diffuse primer dimer band at the bottom of the gel (grey arrowhead).

4.2. Competing processes and PCR optimization

PCR is a competition of several processes, only one of which is the amplification of the desired PCR product. Some of these processes are independent of the presence of a functional template DNA in the reaction. (e.g. the formation and amplification of primer dimers, which are short PCR fragments of unclear origin capable of amplification). Primer dimers can be seen as weak diffuse bands at the bottom of the gel in Figure 2. Because they are short, primer dimers tend to amplify faster than the desired PCR product and once formed are poised to outcompete it. In conventional PCR with high initial template concentration, the desired PCR product may have enough time to fully amplify before primer dimers catch up and take over. As template

concentration is being decreased to achieve limiting dilution, primer dimer gets progressively better chances to dominate the PCR. Thus PCR that works perfectly at high template concentrations may yield nothing but primer dimers at low concentrations of template. The other apparent competing process is some kind of polymerization of PCR products, which probably involves both correct and adverse competitors, which results in generation of “smears” (excessively wide bands on the gel), as well as formation of products that fail to enter the gel. It is thus imperative to design PCR scheme that minimizes primer dimer formation and/or other possible competing processes and maximizes efficiency of the amplification of the correct product. This involves the use of high quality enzyme, good primers and an appropriate thermal profile. We have chosen to fix the reaction buffer composition and thermal profile of the cycle (except for elongation step which is adjusted according with the length of the fragment) and vary only the primers, while keeping their GC content at 50% on average with a decrease towards 3' end to avoid mispriming. In our hands, most primers we design are suitable for smPCR, though one has to be prepared to screen a few combinations of primers if a problem occurs. Such optimization involves running PCR at about 100 template copies per reaction with different primer combinations for different number of cycles and comparing the quality of the product at the optimal number of cycles. We compare the intensity of the main band, the presence/intensity of primer dimers and other undesired effects such as “smearing” of the main product or the presence of additional bands).

4.3. Contamination

Obviously, PCR starting with low concentrations of DNA is more prone to contamination than regular PCR. To avoid contamination, smPCR reactions should be assembled in a clean room, preferably in a biosafety cabinet with filtered air. All reagents and materials used should be specially designated for that purpose. It is important to realize, however, that “contamination” in most cases does not refer to contamination with foreign DNA or the correct PCR product. The most dangerous type of PCR contamination, especially where long smPCR is performed, are the competing adverse PCR products such as those mentioned in section 4.2. These competitors, if they contaminate PCR reactions, will cause the reactions to fail. We were able to achieve stable high quality smPCR only after PCR assembly had been transferred to clean environment.

5. Sample protocol

5.1. Standard protocol

A typical smPCR protocol consists of two stages of nested PCR, though shorter PCR fragments may get well amplified in one stage.

First stage:

Reaction buffer composition:

combine on ice for 100 ul total volume:

Water	to 100ul
10x LA PCR buffer (+Mg) (TaKaRa)	10ul
dNTP mix (TaKaRa)	15ul
External nested primers, 2uM	10ul+10ul
bromphenol blue, 1mg/ml	2ul
DNA template	as needed

mix gently

TaKaRa LA DNA polymerase, 5u/ul

2ul

mix gently,

aliquote into a multiwell PCR plate 10ul per well using an automatic pipette, cover reaction mixture in each well with a droplet of mineral oil.

Typical PCR conditions:

40 cycles, each consisting of a 20 second denaturation step at 94°C, and an elongation step, 16 min (~1 min per Kb) at 68°C.

Second stage:

Once first stage reaction is completed, 1 ul of each of the first stage PCR reactions is transferred into a well of another plate loaded with second stage reaction mixture (10ul per well, reaction composition is the same as first stage except for the use of internal nested primers), and covered mineral oil.

Second stage PCR is typically run for 25 cycles.

5.2. Alternative “droplet” protocol

smPCR experiments routinely use full 96-well plates. Automatic pipetting makes the process very efficient, so that it does not limit the procedure as far as effort is concerned. Automatic pipetting cannot be used during transfer of first stage reactions into the second stage plate. To overcome this difficulty, once PCR works fine in the above describes two-plate arrangement, it is beneficial to switch to a “droplet” protocol, where first stage PCR is performed in 1–3 ul droplets suspended within mineral oil. Everything in the protocol is kept the same except the concentration of polymerase is doubled and mineral oil is dispensed in to wells prior to dispensing reaction mixture. PCR mixture is then dispensed into the oil droplet sitting on the bottom of the well. Bromphenol blue helps to visualize droplets. The main advantage of this protocol is that second stage PCR now can be initiated simply by adding 10ul of second stage PCR mixture into the first-stage plate wells by an automatic pipet and putting the plate into PCR machine for more cycles. The fresh PCR mixture sinks through the layer of mineral oil and mixes with the droplet of the first PCR reaction.

Note on the number of cycles—The number of PCR cycles needed to amplify from a single template DNA copy varies significantly depending on the length of the amplified fragment, the shorter fragments needing less cycles. On the other hand, shorter PCR fragments are less prone to “smearing” and thus shorter PCR can be run for many extra cycles without compromising the reaction. Longer PCR fragments, however, need more precise number of cycles, as they will, once fully amplified, yield to smearing, ultimately resulting in the disappearance of the identifiable band. The border between “short” and “long” fragments is blurred, it may be tentatively put at 6kb.

6. Validations and troubleshooting

6.1. Are these really single molecules?

In principle, the appearance of a clear “yes or no” pattern of multiwell PCR (Figure 1) almost guarantees that each PCR originates from a single molecule. To confirm this, however, it may

be helpful to run a smPCR on a mixture of two types of mtDNA, e.g. a mixture of DNA from two unrelated persons, with enough differences between the two genomes, and if sequencing of the supposed smPCR products yields pure sequences of each type rather than a mixture of both, then single template origin is confirmed.

6.2. Too much vs. too little amplification

In most cases, negative result in smPCR may be described either as too little/no product or too much of an incorrect product.

If there seems to be no product visible in the gel, the first thing to check is the presence of primer dimers (a characteristic diffuse banding close in mobility close to bromophenol electrophoretic dye – see Figure 2 for an example). Primer dimers are visible only when the gel is analyzed soon after completion of electrophoresis and if electrophoresis is not too long (~30 min in 1% agarose) due to rapid diffusion. If primer dimers are absent, a gross error in assembling or running PCR is the primary suspect, e.g. no polymerase, no buffer, faulty PCR machine, etc.

If primer dimers are the only product present, or if expected PCR fragment appears very weak, PCR likely needs optimization. In this situation a positive control PCR, i.e. PCR reaction definitely containing amplifiable templates, e.g. 10 molecules per reaction. If absolutely no sign of expected product is present, but a positive control PCR yields normal PCR products, then DNA concentration may be too low, so that there are no templates to amplify. DNA concentration should be increased until the desired product is visible and PCR parameters optimized (section 4.2). Feedback contamination with primer dimers also may be an issue, thus change of the primers may help.

The other common possibility is that a product is present, but it either looks like a “smear” or as an intense band at the very top, apparently unable to enter the gel. Smearing may be related to excess of DNA polymerase, and decrease 1.5–2-fold, other parameters fixed, is recommended. In some cases these smears may imply low DNA quality or may require a complete redesign of PCR system. Product at the top of the gel usually implies too many cycles and in most cases is an easy to fix problem.

Acknowledgements

This work was supported in part by a grants from NIH/NIA AG019787 and from United Mitochondrial Disease Foundation to KK. PC received a predoctoral fellowship from the Ministry of Education and Science.

References

1. Vermulst M, Bielas JH, Kujoth GC, Ladiges WC, Rabinovitch PS, Prolla TA, Loeb LA. Nat Genet 2007;39:540–3. [PubMed: 17334366]
2. Sato A, Nakada K, Akimoto M, Ishikawa K, Ono T, Shitara H, Yonekawa H, Hayashi J. Proc Natl Acad Sci U S A 2005;102:6057–62. [PubMed: 15829586]
3. Khaidakov M, Chavannes-Turesky N, Cooney CA, Dupont-Versteegden EE, Kennedy RH, Siegel ER, Khaidakova G, Shmookler Reis RJ. Exp Gerontol 2005;40:396–402. [PubMed: 15919591]
4. Khrapko K, Collier H, Andre PC, Li XC, Foret F, Belenky A, Karger BL, Thilly WG. Nucleic Acids Res 1997a;25:685–93. [PubMed: 9016616]
5. Bielas JH, Loeb LA. Nat Methods 2005;2:285–90. [PubMed: 15782221]
6. Zheng W, Khrapko K, Collier HA, Thilly WG, Copeland WC. Mutat Res. 2006
7. Monnat R Jr, Loeb LA. Proc Natl Acad Sci U S A 1985a;82:2895–9. [PubMed: 2986144]
8. Zhang D, Mott JL, Chang SW, Denniger G, Feng Z, Zassenhaus HP. Genomics 2000;69:151–61. [PubMed: 11031098]

9. Khrapko K, Collier HA, Andre PC, Li XC, Hanekamp JS, Thilly WG. Proc Natl Acad Sci U S A 1997b; 94:13798–803. [PubMed: 9391107]
10. Simon DK, Lin MT, Ahn CH, Liu GJ, Gibson GE, Beal MF, Johns DR. Genomics 2001;73:113–6. [PubMed: 11352572]
11. Muller-Hocker J, Seibel P, Schneiderbanger K, Kadenbach B. Virchows Arch A Pathol Anat Histopathol 1993;422:7–15. [PubMed: 7679851]
12. Nekhaeva E, Bodyak ND, Kraytsberg Y, McGrath SB, Van Orsouw NJ, Pluzhnikov A, Wei JY, Vijg J, Khrapko K. Proc Natl Acad Sci U S A 2002;99:5521–26. [PubMed: 11943860]
13. Dubec SJ, Aurora R, Zassenhaus HP. Rejuvenation Res 2008;11:611–9. [PubMed: 18593279]
14. Kraytsberg Y, Khrapko K. Expert Rev Mol Diagn 2005;5:809–15. [PubMed: 16149882]

Figure 2.
An example of a “good” single molecule PCR scenario.