



Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations

Yevgenya Kraysberg and Konstantin Khrapko[†]

A critical review of the clone-by-clone approach to the analysis of complex spectra of somatic mutations is presented. Studying of *a priori*, unknown, somatic mutations requires painstaking analysis of complex mixtures of multiple mutant and non-mutant DNA molecules. If mutant fractions are sufficiently high, these mixtures can be dissected by cloning of individual DNA molecules and scanning of the individual clones for mutations (e.g., by sequencing). Currently, the majority of such cloning is performed using PCR fragments. However, post-PCR cloning may result in various PCR artifacts – PCR errors and jumping PCR – and preferential amplification of certain mutations. This review argues that single-molecule PCR is a simple alternative that promises to evade the disadvantages inherent to post-PCR cloning and enhance mutational analysis in the future.

Expert Rev. Mol. Diagn. 5(5), xxx–xxx (2005)

Somatic mutations are responsible for developmental defects, cancer and possibly aging. To fully appreciate the abundance and impact of somatic mutations, and to study their sources, one needs to measure mutational spectra – the distributions of (*a priori* unknown) mutations present within a certain DNA fragment, by type and frequency.

Analysis of somatic mutations involves separation of complex mixtures

Measuring somatic mutational spectra has always been a challenge, the two main problems being the low frequency of each individual mutation and the high complexity of mixtures of mutations that need to be analyzed. Although overall load of somatic mutations in the genome of a typical cell may be high, the probability of a mutation occurring within the relatively short fragment selected for analysis is usually quite low. Furthermore, a typical mutational spectrum, even for a short DNA fragment, consists of a large number of mutational hotspots of variable intensity [1]. As a result, when studying somatic mutations, one has to be prepared to deal with a complex mixture of a large number of wild-type molecules containing individually rare mutant molecules of multiple types.

Separation of complex wild-type/mutant mixtures is a difficult analytical problem. A few efficient separation methods, mostly based on differential melting of the wild type and the various mutant DNA molecules, have been used for this purpose. However efficient these approaches are, they do not guarantee that all mutations are detected. Furthermore, once mutants are separated from the wild type and from each other, they still need to be further purified for identification of mutations by sequencing. This last purification step is sometimes quite laborious, while separation itself requires either expensive equipment or a lot of skillful effort.

Clone-by-clone mutational scanning:
a tour de force approach in somatic
mutation analysis

Alternatively, a complex mixture of mutations can be analyzed in the molecule-by-molecule mode. In this approach, a large number of molecules from the mixture are cloned, and each molecule is scanned for the presence of a mutation(s), for example, by sequencing in the case of point mutations. The distribution of mutations in the original sample is then approximated by the distribution of mutations among

CONTENTS

Analysis of somatic mutations involves separation of complex mixtures

Clone-by-clone mutational scanning: a tour de force approach in somatic mutation analysis

Post-PCR cloning: simplicity at the cost of PCR artifacts

Single-molecule PCR in mutational analysis: PCR without the artifacts

Expert commentary & five-year view

References

Affiliations

[†] Author for correspondence
Beth Israel Deaconess Medical
Center & Harvard Medical School,
21–27 Burlington Avenue, Boston
MA 02215, USA
Tel.: PHONE
Fax: FAX
khrapko@hms.harvard.edu

KEYWORDS:
aging, allelic preference,
jumping PCR, mitochondrial
DNA, mutational spectra,
PCR errors, single-molecule
PCR, somatic mutations,
recombination

molecules sampled by the cloning procedure [2]. This is a simple straightforward approach where one's ability to detect mutations is limited by the number of clones that can be analyzed. As the cost of DNA sequencing continues to decrease and the high-throughput format becomes routinely available in molecular biology laboratories, the clone-by-clone approach becomes highly competitive. The approach is especially suitable for hypermutable regions of the genome, where the expected fraction of mutant clones is relatively high and thus not too much effort is spent on nonproductive sequencing of the wild-type DNA.

Mitochondrial DNA (mtDNA) is in fact one of the most hypermutable components of the genome. Somatic mutant rates in mtDNA may be approximately 3 orders of magnitude higher than in average nuclear DNA [1], and mutant fractions in the order of one per 10,000 are not unheard of [3]. In other words, one can expect to find one mutation in approximately a dozen sequencing reactions. This price appears to be very attractive, and multiple laboratories use cloning in studies of somatic mutations in mtDNA [4–15]. Some recent studies invested heavily into sequencing of many millions of base pairs of clones in search of mutations [3,16–20]. Clone-by-clone mutational analysis is by no means limited to mtDNA, it can be and is successfully applied whenever one needs to resolve a relatively enriched mixture of mutations, which includes the many applications in toxicology where mutants can be enriched by positive selection in cultured cells [21].

Post-PCR cloning: simplicity at the cost of PCR artifacts

Earlier studies involving clone-by-clone mutational analysis used direct cloning of cellular DNA in bacterial vectors [4,5], or used DNA from cloned mutated bacteria [2] or cells [21]. Recently, however, most researchers choose to clone PCR fragments, rather than the native cellular DNA. This post-PCR cloning procedure is attractive due to its relative simplicity. One does not need to purify mtDNA from nuclear DNA, and commercially available kits facilitate cloning of PCR fragments compared with the rather laborious direct cloning, although it is still necessary to make vector constructs, transfect bacteria and grow/analyze colonies. Furthermore, many experiments, including those performed in single cells or in specific small areas of the tissue, do not provide sufficient amounts of DNA for direct cloning. Finally, while feasible for the low-complexity genome such as mtDNA, direct cloning is much more difficult for nuclear DNA.

Despite its obvious advantages, the post-PCR cloning approach has serious drawbacks. PCR is known to generate artifacts, such as PCR errors and template rearrangements (jumping PCR), and to enrich some mutant alleles at the expense of others (allelic preference). Once the post-PCR mixture of DNA fragments is subject to cloning, PCR artifacts are also cloned and thus become legitimized and indistinguishable from the genuine mutations. If the frequency of cloned artifacts is comparable to or exceeds the expected fraction of genuine somatic mutations, then mutational analysis is compromised. The different sources of PCR artifacts and their impact on mutational analysis by post-PCR cloning will now be considered.

PCR errors do interfere with mutation analysis: the mtDNA example

It has been recognized for some time that DNA polymerases used in PCR are error prone and thus, PCR product contains random artificial mutations. Whether PCR errors represent a true problem in mutational analysis depends on how low a mutant fraction one intends to measure. The reports that used the (error-free) direct cloning approach allow estimation of the expected mutational rates in the mitochondrial genome, a likely candidate for post-PCR, cloning-based, mutational analysis. For human mtDNA, studies reported approximately three mutations per 10^5 nucleotides in peripheral lymphocytes [4], in leukemic cells of middle aged donors [22], and the retina of a 71-year-old donor [5]. The authors observed approximately 2.5 mutations per 10^5 nucleotides in a 40-year-old muscle [13]. A very large sequencing study reported approximately 2.6 mutations per 10^5 nucleotides in cultured human cells, and one mutation per 10^5 nucleotides in muscle and 0.4 mutations per 10^5 nucleotides in brain of 6-month-old mice [19]. In conclusion, to analyze somatic mutations in mtDNA, one has to be prepared to detect approximately one mutation per 10^5 bp.

Reported error rates of thermostable polymerases commonly used in PCR vary widely from 2×10^{-4} errors/bp/duplication for Taq polymerase to 6×10^{-7} errors/bp/duplication for pfu polymerase [23]. Assuming that a typical PCR amplification involved approximately 20 duplications (10^6 -fold amplification), these rates translate into a range of error fractions from 0.5 to 200 errors per 10^5 nucleotides. It therefore appears that even most high-fidelity polymerases are not quite suitable for post-PCR cloning since the fraction of errors is comparable to the *in vivo* mutant fraction. For example, Cantuti-Castelvetri and coworkers reported six to nine errors (depending on clonality assumptions) per 10^5 nucleotides with the HF2 high-fidelity polymerase from Clontech after approximately 30 duplications [18]. Pfu polymerase was reported to provide better results – approximately 1.4 errors per 10^5 nucleotides after 25 cycles of PCR [3]. In conclusion, fractions of PCR errors observed in post-PCR cloning procedures are comparable to the fractions of *in vivo* mutations and thus may interfere with mutational analysis.

Jumping PCR: a problem for preserving associations between mutations & for the analysis of recombinants

PCR amplification may involve template jumping [24], a condition where DNA polymerase switches from one template to another while maintaining synthesis of a continuous nascent DNA strand. The result is a composite molecule originating from two templates. Switching of templates has no net effect if the templates are identical or differ by one mutation. However, if the two templates differ by at least two mutations located on each side of the switching point, jumping produces a new type of molecule identical to a product of recombination between the two templates. In somatic mutational analysis, most molecules in a mixture are assumed to be wild type, and the probability of two mutant molecules meeting is low. If

two mutant molecules do meet and template jumping does create a double mutant molecule, then there is a high probability that the two mutants will be dissociated on the next encounter with a wild-type template. This logic seems to indicate that template switching is not a problem in mutational analysis, but in fact it is.

The jumping PCR artifact becomes important if there are more than one mutation per molecule within the DNA region under consideration. As the DNA fragments that are being scanned for mutations increase in length (it is not uncommon to scan the whole 16-kb mitochondrial genome [25,19]), the probability of finding a multiple-mutant genome becomes quite substantial. In this case, template switching may separate mutations that were originally on the same DNA molecule or, conversely, join originally disjointed mutations on the same DNA strand. This may create combinations of mutations that may not have existed in the original sample, but will be indistinguishable from the real ones upon cloning.

The information regarding mutation association may seem irrelevant to someone convinced that mutagenesis is a completely random process. However, preliminary data indicate that mutations are nonrandomly distributed between mtDNA molecules [KRAYTSBERG & COWORKERS, UNPUBLISHED DATA]. Certain mtDNA molecules appear to contain multiple mutations. This may indicate that there are mutator cells or mutator mitochondria within cells. Alternatively, certain combinations of molecules may have an advantage in propagation of the mtDNA molecules that harbor them. Information regarding such nonrandom associations would have been erased by jumping PCR.

Obviously, jumping artifact is unacceptable in any experiments studying DNA recombination, since products of jumping PCR, once they are cloned, are indistinguishable from genuine recombinant DNA molecules. A clear example of abundant artifactual recombinants created by precloning PCR has been reported recently [19].

Allelic preference: a problem in quantification of deletion mutations

Allelic preference arises when templates with sequence variants present in a mixture of DNA fragments subject to PCR are amplified with unequal efficiencies. In a mixture of mutations, any mutation with an increased amplification efficiency will be over-represented in the post-PCR mixture. Allelic preference becomes particularly severe in the case of deletion mutations. Mutant DNA with large deletions may be amplified with drastically higher efficiency than the wild-type molecules due to its shorter length. Allelic preference is a long-standing problem in quantification of mtDNA deletions that are involved in mitochondrial disease and likely play a role in human aging [26]. Not only are deleted mtDNA better amplified than the wild type, the extent of allelic preference is highly sensitive to slight changes in PCR conditions, so not only do PCR-amplified mixtures deliver biased results, the comparison between different laboratories and even different PCR reactions is difficult.

Single-molecule PCR in mutational analysis:

PCR without the artifacts

The authors have argued so far that although direct cloning in bacterial vectors is an excellent approach in mutational analysis, it is laborious and in many cases unfeasible. Post-PCR cloning is a much simpler approach but is subject to the various PCR artifacts. Therefore, the authors propose that single-molecule PCR (smPCR), an approach already used in a number of important applications, should be used in mutational analysis in place of post-PCR cloning. smPCR is even simpler than post-PCR cloning and does not suffer from many of its drawbacks.

It was demonstrated over a decade ago that PCR can be efficiently performed on a single DNA template [27]. This is achieved by performing multiple PCRs at limiting dilution, where DNA concentration is so low that many of the reactions (usually ~50%) by pure chance do not receive any template molecules at all and thus produce no PCR product. Under such conditions, the positive reactions are most likely to have been initiated by a single template molecule. Since its discovery, smPCR was used for mutational analysis of microsatellites [28] and is currently used for studies of recombination [29,30]. The method was reinvented several times under different names and used for construction of expression libraries [31] and for analysis of cancers for known mutations (digital PCR) [32,33]. Recently, smPCR was used to generate colonies – spots of PCR products originating from a single template that are immobilized in a thin gel layer on glass surface, similar to bacterial colonies in agar [34].

PCR amplification of a single molecule results in a clone of PCR fragments. Thus, using smPCR, one can bypass the cloning step in post-PCR cloning in mutational analysis, thereby simplifying the process. More importantly, it appears that when performed in the single-molecule mode, PCR amplification is immune to PCR errors, template jumping and allelic preference. This makes smPCR particularly suitable for mutational analysis of unknown somatic mutations, which, as shown above, suffers from all these artifacts. Therefore, it is a pity that smPCR has not yet established itself in the field of somatic mutation research. The methodological aspects of smPCR as applied to somatic mutation analysis have been described in a recent book chapter [13]. The reasons why smPCR is (or can be made) resistant to the various PCR artifacts are now discussed.

smPCR reduces PCR error rate & permits filtering of the errors

One of the advantages of the digital data format is the availability of error-correction filtering, and digital output of smPCR is no exception. The principle of smPCR error correction is described in detail in FIGURE 1. Ideally, in PCR started from a single double-stranded DNA template, PCR errors cannot be present in more than 25% of the resulting PCR fragments (FIGURE 1). Indeed, an error can be introduced only during replication of an existing strand, and at the end of the first PCR cycle there will be four single strands and only one mutation (hence 25%). Thus, PCR errors can be easily identified and

