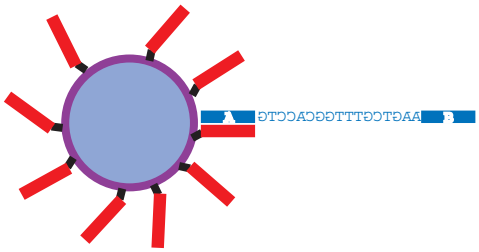
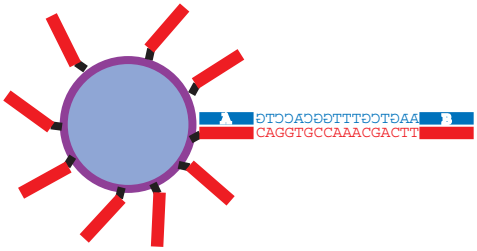
Figure S1: Summary of the emulsion PCR (emPCR) process

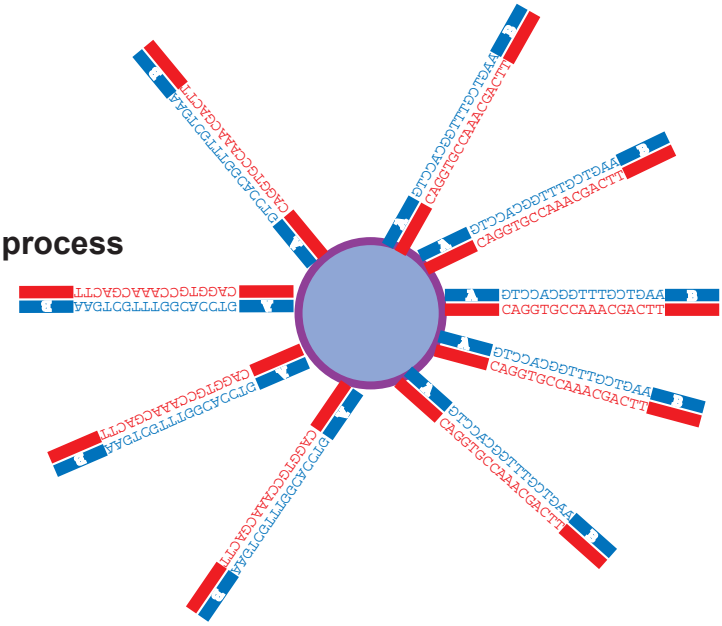
1b) Subsequent ligation and purification steps leave original, single-stranded template molecules bound to ligator sequences 'A' and 'B'

1c) These single-stranded molecules are captured by oligonucleotide probes complementary to ligator 'A' that are bound to the emPCR beads.

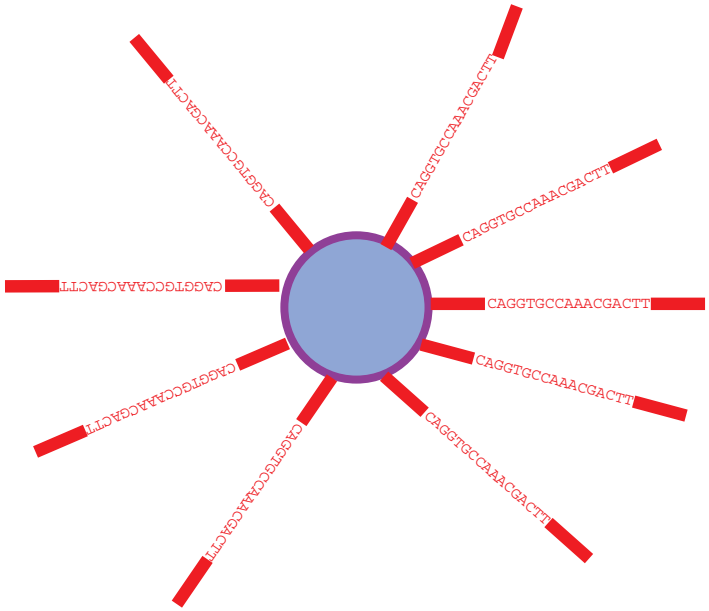




1d) During the first round of emPCR, a complementary molecule to the captured original template molecule is synthesised from the probe



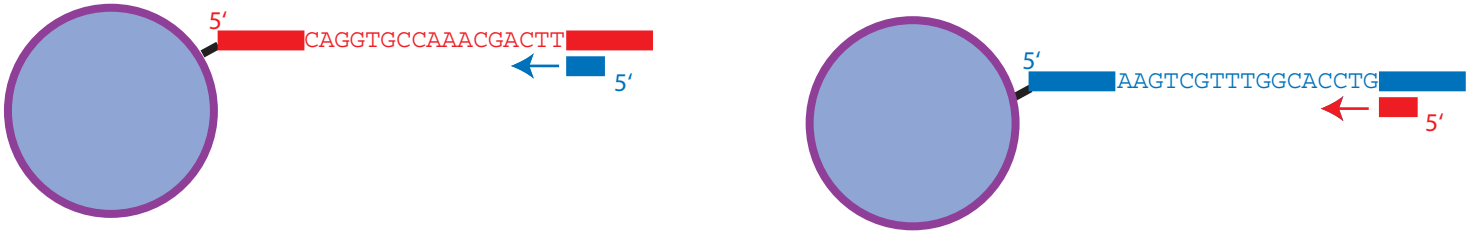
1e) Decendent molecules from the emPCR process are bound to the emPCR bead by the remaining probes



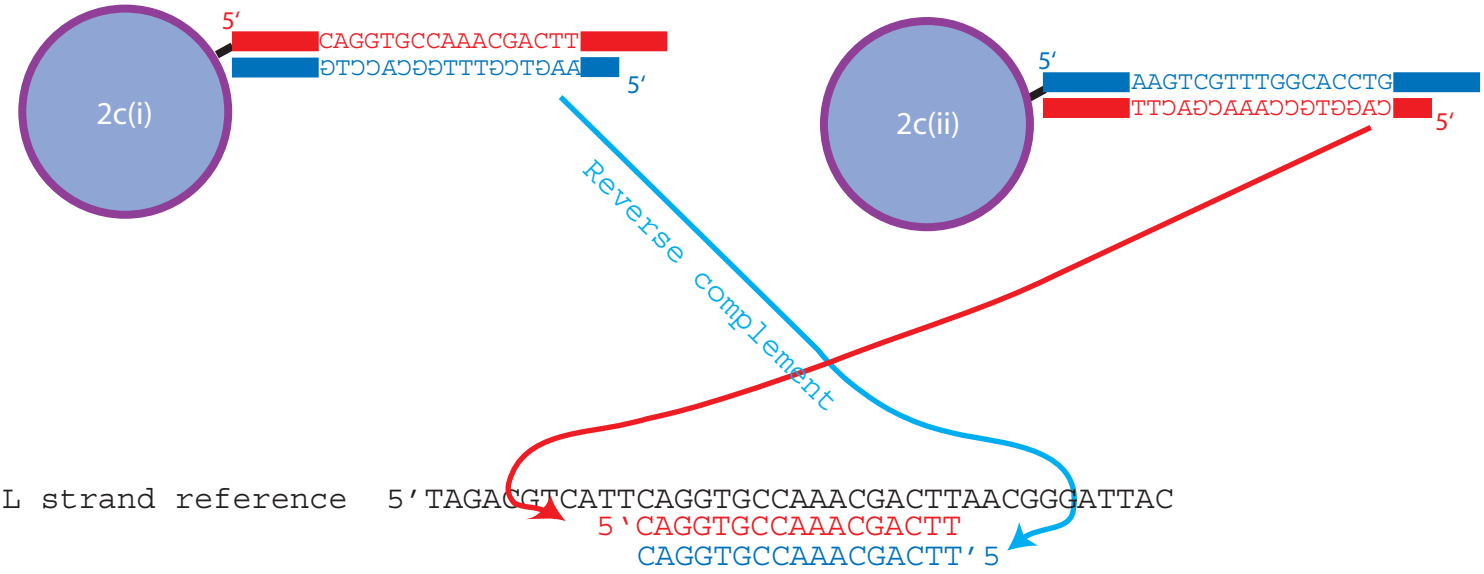
1f) The captured DNA fragments dissasociate during further processing steps. Only the captured-single stranded molecules, which are the complement of the original single-stranded template molecule remain. These form the template for subsequent pyrosequencing (See Fig. S2)

Fig. S2: Pyrosequencing of captured emPCR products and subsequent identification of strand of origin of the original template molecule

2a) Following emPCR captured single-stranded DNA undergoes pyrosequencing using primers that are complementary to the 3' ligator sequence



2b) Generated sequence is the reverse complement of the captured single-stranded DNA molecule



2c) When aligning the sequence to a reference sequence (here a mitochondrial L strand):

2c(i) Sequence generated from a captured L (red) strand must be reverse complemented for alignment

2c(ii) Sequences generated from a captured H (blue) strand can be directly aligned to the reference sequence

2d) Therefore: During data analysis the original orientation of the template molecule can be unambigously determined by assessing whether the sequence can be directly aligned or not

2e) Furthermore, as the captured single-stranded DNA molecules are the complement of original single-stranded molecules (see Fig. 1):

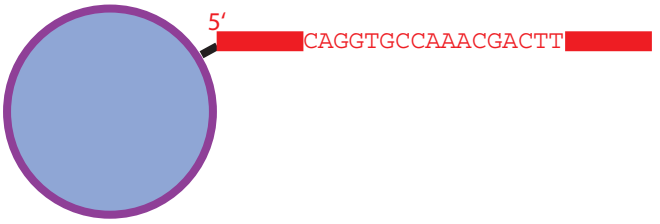
2e(i) Sequence generated from a captured L (red) strand ultimately originate from an original single-stranded L strand molecule

2e(ii) Sequence generated from a captued H (blue) strand ultimately originate from an original Single-stranded H strand molecule

2f) This statement can be generalised to state: DNA sequences generated through sequencing-by-sythesis directly describe the original single-stranded template molecule

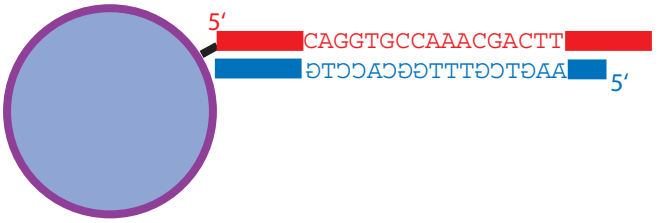
Figure S3: Identification of DNA miscoding lesions in sequence data

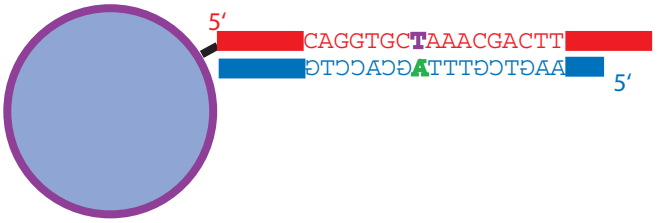
3a) Captured single-stranded DNA molecule post emPCR process (as Fig. S2a)



3b) Captured single-stranded DNA molecule containing C→T miscoding lesion post emPCR process. The captured molecule is the complement of an original single-stranded molecule.Therefore the observed miscoding lesion derives from an original G→A damage event.

3c) Captured molecules undergo pyrosequencing (see Fig. S2). A G→A change is observed when the sequence derived from the damaged molecule (Fig. S3b) is compared to the sequence derived from the undamaged (Fig. S3a) molecule.





3d) As the generated sequence directly describes the original single-stranded template molecule (Fig. S2f), the observed G→A miscoding lesion accurately describes the original G→A damage event.