© 2007 Nature Publishing Group http://www.nature.com/naturemethods

MegaPlex PCR: a strategy for multiplex amplification

Linda Strömqvist Meuzelaar[1], Owen Lancaster[1], J Paul Pasche[2], Guido Kopal[2] & Anthony J Brookes[1]

'MegaPlex PCR' is a robust technology for highly multiplexed amplification of specific DNA sequences. It uses target-specific pairs of PCR primers that are physically separated by surface immobilization. Initial surface-based amplification cycles are then coupled to efficient solution-phase PCR using one common primer pair. We demonstrate this method by co-amplifying and genotyping 75 unselected human single-nucleotide polymorphism (SNP) loci.

The key challenge in developing a high-multiplex amplification procedure is preventing excessive off-target priming by the many primers in the reaction[1–3]. Several previously developed methods use surface immobilization of primer pairs as a way to separate reactions and thus limit primer-dimer formation[4–6], but drawbacks of these methods include inefficient amplification, loss of reactants from the surface and considerable primer-dimer formation within pairs of primers. MegaPlex PCR likewise uses physical separation of primer pairs to prevent their interaction, but it additionally includes 'common sequences' at the 5′ ends of the surface primers so that early-stage amplicons can be moved into the solution phase for co-amplification by highly efficient and unbiased PCR using a single common primer pair[6–8] (**Fig. 1**). To reduce primer-dimer formation, MegaPlex PCR uses surface primers that are made partially double-stranded by base-pairing their common 5′ sequences with complementary 'barrier oligonucleotides'. A detailed description is available in **Supplementary Methods** online.

During the development of MegaPlex PCR we evaluated a range of targets, input DNAs and reaction conditions. We used membrane arrays and microbeads as binding surfaces, with the beads now routinely attached to microtiter plate wells to provide a platform compatible with automation. Using the microbead support we found MegaPlex PCR to be effective with as little as 200 ng of human genomic DNA, and in various 15-plex reactions it recovered many different target sequences of up to at least 500 bp in size, with little bias against larger amplicons (**Supplementary Fig. 1** online). As occasional targets (~1/10 in preliminary studies) generated severe primer-dimer artifacts, regardless of the multiplex level of the experiment, optimization of MegaPlex PCR was required. One option was to exclude these targets by empirical or computational prefiltering, but we selected a modification that avoided the need for target preselection.

This modification involves enriching the input genomic DNA for sequences of interest via a crude solution-phase multiplex PCR using specific primers for all the amplification targets. The primer sequences for this can match the specific portions of those on the solid surface, or they can be designed to prime slightly outside these sites in the genomic DNA. This strategy is used routinely in digital

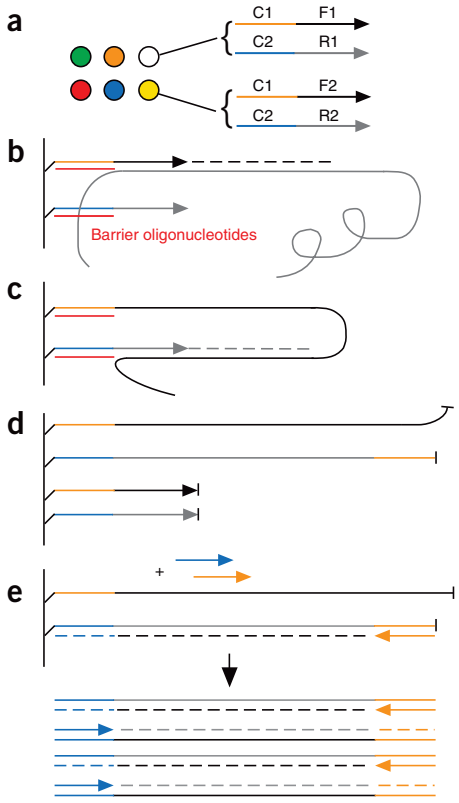

Figure 1 | Schematic representation of MegaPlex PCR. (**a**) Colored circles indicate surface features with different primer pairs, with common 5′-end sequences (C1 and C2) and target-specific 3′-end sequences (F1 and R1, F2 and R2). (**b**) In the first cycle of solid-phase PCR single-stranded target DNA is hybridized and specifically primed by surface-bound primers (themselves duplexed with 'barrier oligonucleotides'), which are then extended to copy the target sequence. Template DNA is then removed. (**c**) In the second cycle of solid-phase PCR the first extension product hybridizes to the other primer and a target with common primer regions at both ends is produced. (**d**) Surface-bound molecules are blocked at their 3′ ends (vertical solid bars) to prevent subsequent priming. (**e**) Common primers are added for standard PCR.

[1]Department of Genetics, University of Leicester, University Road, Leicester, LE1 7RH, UK. [2]Roche Applied Science, Nonnenwald 2, 82372 Penzberg, Germany. Correspondence should be addressed to A.J.B. (ajb97@leicester.ac.uk).

RECEIVED 28 JUNE; ACCEPTED 23 AUGUST; PUBLISHED ONLINE 16 SEPTEMBER 2007; DOI:10.1038/NMETH1091

Tecan Ex. 2136, Page 1 of 3
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

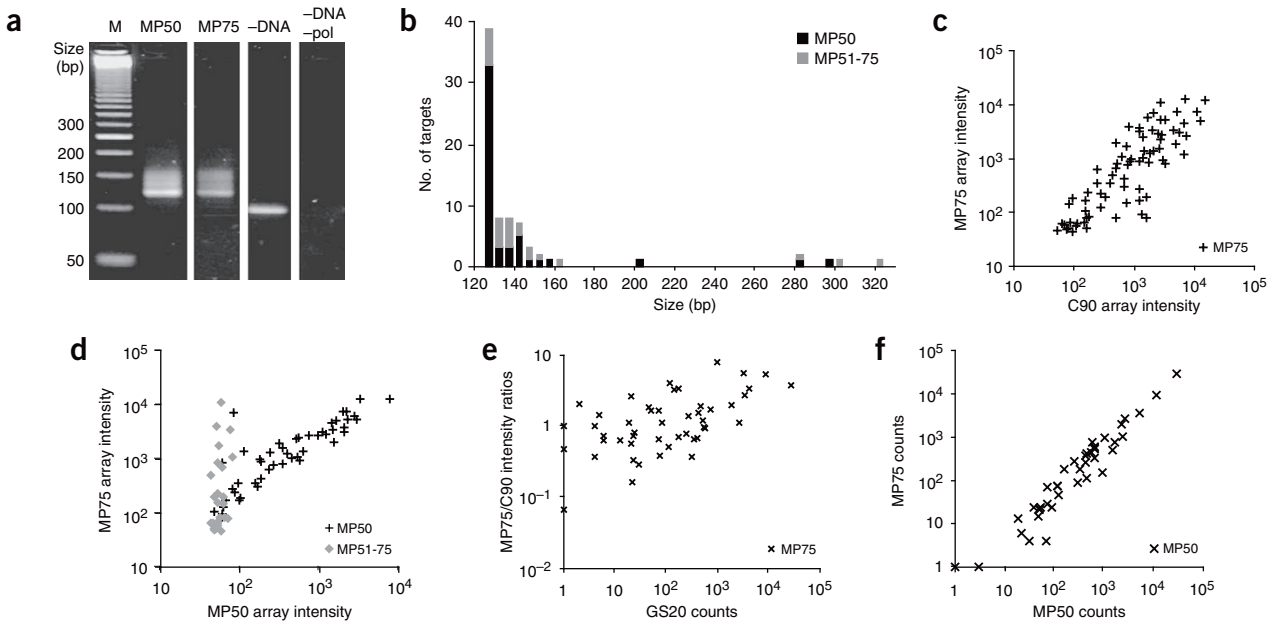

Figure 2 | Analysis of the MP50 and MP75 amplification products. (**a**) MegaPlex PCR products on a 3% agarose gel. Control reactions included the same primers as the MP50 reaction and show artifacts owing to cross-priming between surface primers that are absent from the other reactions. (**b**) Expected size distribution of the full set of MP50 targets and the 25 additional targets (MP51-75) only present in MP75. Given sizes include the added common primer regions, shown in bins of 5 bp. (**c**,**d**) Microarray analysis. Array hybridization signals are plotted per target. Median array feature signals for the MP75 amplification, plotted against the C90 control data (**c**). MP75 signals against the equivalent data for MP50 (**d**). (**e**) Relative recovery rates for the MP75 targets, with estimates from the GS20 sequencing analysis on the *x* axis against estimates from the microarray analysis on the *y*-axis. Recovery rate estimates from the microarray data were derived by calculating the signal ratio between MegaPlex product signals and their paired C90 signals after subtracting nonspecific background. Recovery rate estimates from the GS20 data were derived by counting the number of times each target sequence was observed. (**f**) A comparison of GS20 counts for MP75 and MP50.

© 2007 Nature Publishing Group http://www.nature.com/naturemethods

molecular counting and genome mapping studies that use limiting-dilution samples, with multiplexing up to at least 1,200 (ref. 9). Although this strategy is less specific than conventional PCR, it can be expected to efficiently enrich ~80% of all target sequences for which primers are included (Paul Dear, Medical Research Council, Laboratory of Molecular Biology; personal communication). The increased abundance of these fragments allows them to outcompete artifacts resulting from primer-dimer formation. An additional benefit of first enriching the input DNA is that less genomic DNA is needed per experiment (tens versus hundreds to thousands of nanograms). Other multiplex amplification methods have also used such an enrichment step[10].

To demonstrate the complete MegaPlex PCR procedure, we used it to amplify and genotype 75 randomly selected prevalidated human SNPs (**Supplementary Table 1** online). To assess technical reproducibility, we amplified all 75 targets (MP75) as well as a subset of 50 loci (MP50). The protocol used microbeads adhered to the walls of microtiter plate wells as the solid surface, and genomic DNA samples enriched for the MP75 and the MP50 sets of targets. This level of multiplexing of completely unselected sequences would not work well using standard multiplex PCR in solution, as demonstrated by the nonspecific product smear generated by the initial DNA enrichment in these experiments (**Supplementary Fig. 2** online). More sophisticated strategies for solution-phase multiplex PCR[11,12] can operate at this multiplex level, but they require careful preselection of targets.

We first examined the MP50 and MP75 Megaplex PCR products by agarose gel electrophoresis (**Fig. 2a**). This revealed amplified material that closely matched the expected size range for the two sets of targets (**Fig. 2b**), with larger amplicons being too faint to see because we included few such targets in the experiment. Notably, we detected no primer-dimers in this preliminary analysis. Then we conducted secondary singleplex PCRs on aliquots of the MP50 product for a random set of 21 of the loci, of which 20 produced clean gel bands of uniform intensity and of the expected sizes (data not shown). These initial results thus implied that the majority of the targeted SNP loci were effectively amplified by MegaPlex PCR.

To further assess the MP50 and MP75 products we used (i) microarray hybridization on the GENIOM platform[13], and (ii) single-molecule counting by sequencing on the Roche Genome Sequencer 20 system (GS20)[14].

For the hybridization analysis, we used oligonucleotide arrays carrying several sequence representations of each of the 75 SNP loci and of a control set of 42 unrelated single-copy human genome fragments, all of which were duplicated on the array. We hybridized biotin-labeled MP50 and MP75 products onto these arrays, allowing subsequent fluorescence detection with streptavidin–R-phycoerythrin conjugate. To control for differential hybridization efficiency among the 75 targets, we performed a control hybridization with a normalized pool (called C90) of all 75 MegaPlex targets plus 15 additional DNA fragments that were represented on the array but not present in the MegaPlex amplifications. Upon comparing the MP50 and the MP75 signals to those of the control C90 hybridization, we observed uniform target recovery rates in each experiment (**Fig. 2c** and **Supplementary Fig. 3** online), and excellent reproducibility for targets shared between MP50 and

Tecan Ex. 2136, Page 2 of 3
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

MP75 (**Fig. 2d**). We determined the target-specific relative recovery rates by dividing the MegaPlex product signals by their paired C90 signals. From this analysis we omitted targets with signals below background in the C90 data (6 targets in MP50 and 12 in MP75) as we presumed they were undetectable with the array conditions used. Only four targets were not recovered according to this analysis, as they gave signals weaker than background in the MegaPlex PCR hybridizations even though they were detectable in the C90 hybridization. Among the remaining (quantifiable) targets, the microarray analysis indicated that the great majority of targets were recovered with efficiencies that varied over little more than a tenfold range (**Fig. 2e** and **Supplementary Fig. 3**).

To conduct the GS20 sequence-based analysis we first re-amplified the MP50 and MP75 products by PCR using 'fusion' primers (MegaPlex common primers flanked at their 5′ ends with GS20 primer sequences) to introduce priming sites required by this sequencing technology. We then re-amplified these products by emulsion PCR to produce single-molecule clones on beads, of which around 80,000 were sequenced per experiment. These additional amplification cycles (especially the 'fusion' PCR) could potentially distort individual target recovery levels relative to their representations in the primary MegaPlex products. Nevertheless, counts of each target sequence indicated that most targets were recovered within about a 100-fold range (**Fig. 2e** and **Supplementary Fig. 3**), with excellent reproducibility between experiments (**Fig. 2f**). Only 7 and 15 expected SNP loci were not observed at all in the sequencing reads for MP50 and MP75, respectively. This is consistent with our expectation that up to 20% of loci might not be enriched by the particular genomic DNA preprocessing step we used. If 100% of targets had to be recovered, or the desired multiplex level was considerably greater than that supported by the particular input DNA enrichment method we used (that is, 10^3 or so), one might use an alternative enrichment procedure or empirically preselect for primers that do not generate excessive primer-dimers.

By directly comparing the GS20-based and the array-based data (**Fig. 2e** and **Supplementary Fig. 3**) we found that the relative recovery rates for specific targets estimated by these two analyses were in good agreement, thereby cross-validating these two analytical strategies. We considered the larger overall spread of recovery rates in the GS20 data to be an artifact resulting from the multiple rounds of PCR we had to perform to apply that technology.

Additional analysis of the GS20 data allowed us to determine SNP genotypes for the input DNA sample, as many representations of each chromosome copy (allele) per SNP had been sequenced. If we observed only one allele repeatedly, we scored the sample as homozygous for that allele, whereas if we observed many examples of both alleles, we scored it as heterozygous. The MP50 and MP75 experiments gave consistent genotype indications for targets they had in common (**Supplementary Table 2** online), and we then validated all of these genotype assignments by dynamic allele-specific hybridization[15]. Across 12 heterozygous SNPs the median abundance for the least represented allele was 37.7% (minimum = 15.3%), and across 26 homozygous SNPs the median abundance for the correct allele was 99.6% (minimum = 95.4%).

Finally, the GS20 data revealed that, even without preselecting targets or primer sequences in any way, less than 10% of all the amplified products were primer-dimer artifacts. Almost all of these involved just 3 structures (each at ~3%) wherein dimerization had occurred between highly complementary 3′ ends of primers designed for different targets. Encouragingly, each of these troublesome primers still produced authentic targets. This is a major improvement over previous multiplex PCR strategies, and may allow MegaPlex PCR to support extremely high multiplex applications, such as the direct isolation of many different exons and/or promoters from a complex genome.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS
This study was supported by the EU Sixth Framework Programme integrated project MOLTOOLS (to A.J.B.).

AUTHOR CONTRIBUTIONS
L.S.M. and A.J.B. designed and planned the study. L.S.M. performed the majority of the laboratory experiments. J.P.P. produced the GS20 sequences. G.K., O.L. and A.J.B. performed data analysis of the GS20 data. L.S.M. and A.J.B. prepared the manuscript.

COMPETING INTERESTS STATEMENT
The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Markoulatos, P., Siafakas, N. & Moncany, M. *J. Clin. Lab. Anal.* **16**, 47–51 (2002).
2. Vallone, P.M. & Butler, J.M. *Biotechniques* **37**, 226–231 (2004).
3. Wang, H.Y. *et al. Genome Res.* **15**, 276–283 (2005).
4. Adessi, C. *et al. Nucleic Acids Res.* **28**, e87 (2000).
5. Shapero, M.H., Leuther, K.K., Nguyen, A., Scott, M. & Jones, K.W. *Genome Res.* **11**, 1926–1934 (2001).
6. Pemov, A., Modi, H., Chandler, D.P. & Bavykin, S. *Nucleic Acids Res.* **33**, e11 (2005).
7. Brownie, J. *et al. Nucleic Acids Res.* **25**, 3235–3241 (1997).
8. Shapero, M.H. *et al. Nucleic Acids Res.* **32**, e181 (2004).
9. Eichinger, L. *et al. Nature* **435**, 43–57 (2005).
10. Fredriksson, S. *et al. Nucleic Acids Res.* **35**, e47 (2007).
11. Hansen, A. *et al. Nat. Methods* **4**, 35–37 (2007).
12. Wang, D.G. *et al. Science* **280**, 1077–1082 (1998).
13. Baum, M. *et al. Nucleic Acids Res.* **31**, e151 (2003).
14. Margulies, M. *et al. Nature* **437**, 376–380 (2005).
15. Howell, W.M., Jobs, M., Gyllensten, U. & Brookes, A.J. *Nat. Biotechnol.* **17**, 87–88 (1999).

© 2007 Nature Publishing Group http://www.nature.com/naturemethods

Tecan Ex. 2136, Page 3 of 3
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029