Product length, dye choice, and detection chemistry in the bead-emulsion amplification of millions of single DNA molecules in parallel

Irene Tiemann-Boege[1,3], Christina Curtis[2], Deepali N. Shinde[3], Daniel B Goodman[2], Simon Tavaré[2,3] & Norman Arnheim[3]*

[1] Institute of Biophysics, Johannes Kepler University, Linz, 4040 Austria

[2] Computational Biology Group, Department of Oncology, University of Cambridge, Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, CB2 0RE England

[3] Molecular and Computational Biology Program, University of Southern California, Los Angeles, CA, 90089, USA

SUPPORTING INFORMATION CONTENTS.

Pre-amplification of samples. Approximately 10ngs of human genomic DNA (Clonetech and Coriell Cell Depositories) were amplified in 50µl with 1x Platinum HiFi buffer, 1.25mM $MgCl_2$, 200µM dNTPs each, 0.2µM Forward and Reverse primers and 1U of HiFi Platinum Taq (Invitrogen). Cycling conditions were 2 minutes at 94°C followed by 35 cycles at 95°C for 15 seconds and 68°C for 75 seconds. The product was purified with the QIAquick PCR purification kit (Qiagen) and quantified using a Nanodrop 1000 (ThermoFisher Scientific) spectrophotometer. The amplified material was diluted in water only immediately before use for the emulsion PCR.

Emulsion PCR. Emulsion PCR was performed as described by [1, 2] with the following modifications. Magnetic streptavidin coated Dynabeads M-270 (Invitrogen) were washed and coupled to a reverse dual-biotinylated primer (sequence details are found further on). A total of ~$3x10^6$ amplicons were hybridized to the complementary sequence of the dual biotinylated primer of ~10^7 beads in 20mM Tris-HCl (ph 8.4), 50mM KCl, and 15mM $MgCl_2$ at 94°C for 2 minutes followed by 15 minutes at 68°C with periodical stirring. The liquid was removed using a magnet (Dynal MPC-S). The beads were washed once with water and resuspended in 150µl of 1x Platinum Taq buffer, 8mM $MgCl_2$, 1mM dNTPs (Roche), 9µM forward primer, 0.05µM reverse primer (underlined sequence of the dual biotinylated primer described further on), and 45U of Platinum Taq. Beads were resuspended for at least 5 minutes in this aqueous phase followed by thorough mixing. During this time, the oil phase was prepared with 145mg Abil WE09 (Evonik Degussa), 740µl TEGOSOFT (Evonik Degussa) and 200µl light mineral oil (Sigma). The components were mixed with a vortex and centrifuged for a few seconds. The emulsion was created by adding 600µl of the oil-phase using positive displacement pipetting and a 5mm steel bead (Qiagen) to the 150µl aqueous phase contained in a 2ml cryovial. The cryovial was assembled into a Tissuelyzer holder (Qiagen) and mixed with the Tissuelyzer at 15Hz for 10 seconds and 17Hz for 7 seconds. The emulsion was aliquoted out into 80µl volumes and the PCR was carried out with an initial denaturation step of 2 minutes at 94°C followed by 55 cycles at 95°C for 15 seconds and 68°C for 75 seconds.

Washing and arraying the beads for analysis. This protocol is a modification of published work [1, 2]. The aliquoted emulsion PCR reactions were pooled in a 2ml Eppendorf tube and the beads were cleaned by adding 500µl of isopropanol, vortexed for at least 30 seconds at full speed and then centrifuged at 13000rpm. The supernatant was removed and the washing step repeated with 900µl isopropanol. At this point, the beads form a hard pellet. The supernatant was decanted and the beads resuspended in 500µl wash buffer (10 mM Tris-HCl (pH 7.5), 0.1% Triton-X 100, 100 mM NaCl, 1 mM EDTA) 1 by vortexing and pipetting. The waste was drawn off using a MPC magnet. This wash step was repeated until no white slurry could be observed. The beads were washed twice with TE and incubated in 200µl freshly made 0.1M NaOH at room temperature for 5 minutes. The denaturation of the DNA on the beads was finalized by two more washes with NaOH followed by two washes with TE. The beads were poured in 1.5µl of 5.5% acrylamide solution prepared with Protogel 37/1 (National Diagnostics), Rhinohide Gel strengthener (Invitrogen), 0.05% APS and 0.5% TEMED on a silane treated microscope slide (protocol modified from [2]). A monolayer of beads was achieved by placing a 10mm diameter coverslip and the weight of several microscope glass slides on top of the beads. The coverslip was immediately removed after polymerization of the acrylamide.

Ligation of Alexa labeled probes. Ligations were carried out as described in [2] with a few modifications. In short, 100µM anchor primer was incubated with the beads in a ligation hybridization solution (5 mM Tris-HCl 7.5, 1mM $MgCl_2$) for 5 minutes at 56°C and 2 minutes at 42°C. The excess anchor probe was washed with 1E buffer (see above) using a MPC magnet. The beads were separated in 4 different aliquots and a different ligation probe (sequence presented further on) was added at a final concentration of 4µM. The ligation reaction was carried out with 6000U of T4 ligase (NEB) and 1x Buffer 2 (NEB) in a total volume of 50µl at room temperature for 15 minutes and terminated by heating the reaction for 5 minutes at 60°C. Beads were washed with 1E buffer and TE. The ligation and anchor probes used are the following:

Ligation probes (Integrated DNA Technologies):

SNP2 (C) -2C3	/Alexa647/N/SpC3/SpC3/tcgcccacgc
SNP2 (C) -SP9	/Alexa647/N/Sp9/tcgcccacgc
SNP2 (C) no spacer	/Alexa647/tcgcccacgc
SNP2 (C)Cy5	/Cy5/tcgcccacgc

Anchor probe:

SNP2 anchor probe PO_4-GGCGCCAACCTGCCC

Image analysis. Automated image analysis was performed using custom Matlab scripts, which register, segment, and extract the centroids of each bead from the brightfield image. The centroids were used to define the fluorescent objects. Only objects that exceeded a threshold of 2 standard deviations greater than the mode of the pixel intensities of the image (mode-thresholding) were considered for the analysis. Fluorescent objects were then filtered based on shape and size criteria. After applying a conservative background correction, based on morphological opening, shown to perform well on microarray data [3], the average of the 90^{th} percentile of pixel intensities of an object was taken as the foreground for each fluorescent bead. Intensity values were obtained in each channel for all thresholded objects, enabling comparisons between channels. For the analysis of the product length and Alexa crosstalk experiments, the most intense beads were selected by taking the 80^{th} percentile of bead intensities for the channel of interest. Throughout they are referred to as beads with a product (positive beads). For the density plots we considered all the beads obtained after thresholding.

For the product length validation experiment, the sample derived from a double heterozygous individual. Here, quantile normalization was used to adjust the intensity distribution for the two sets of dyes labeling SNP1 and SNP2 since similar distributions are expected for the two dyes querying each SNP. The 80^{th} percentile of bead intensities for a given dye was selected as a conservative threshold and beads exceeding this cutoff in a single channel were called accordingly as “positive”. Beads exceeding the threshold to assign the allelic state of the SNP in both the channels were deemed to contain multi-

templates and were eliminated, unless they were four-fold more intense in one channel (in which case a call was made), whereas beads below the cutoffs were termed empty.

Primers and probe information.

Primers for different product lengths:

FORWARD 66bp: 5' GGCGGGCAGTGTGTATGCAG 3'
FORWARD 175bp: 5' CTCTGGGCCAGGGGCATCCAT 3'
FORWARD 450bp: 5' TGGCCCTGTGCCCAGTGTG 3'
FORWARD 982bp: 5' TCTGCGCTAACCCGCATGCT 3'
REVERSE SNP1: 5' CCACCACCAGGATGAACAGGAAG 3';
REVERSE Dual biotinylated primer (Operon): The underlined sequence is a complementary to the pre-amplified template
5' [bioteg][bio-on]CAGAGCGTCACAGCCGCCACCACCAGGATGAACAGGAAG3'

Primers used to produce the 459bp template:

FORWARD 66bp: 5' GGCGGGCAGTGTGTATGCAG 3',
REVERSE SNP2: 5' GGTCAGCAGGGGCAGGTTGG 3'
5' [bioteg][bioon]CTCCAGGGACACCTGCTTGGGTCAGCAGGGGCAGGTTGG 3'

Allele specific probes (Integrated DNA Technologies). Asterisks denote phosphorothioate bonds:

SNP1 (G) A532 /5Alexa532/CAGGCATCCTCAGC*T* A*C*G (GC: 68%)

SNP1 (A) A594 /5Alexa594/CAGGCATCCTCAGC*T* A*C*A

SNP2 (T)-A488 /5Alexa488/CCCTGTCGCCCA*C*G*T (GC: 78.6%)

SNP2 (C)-A647 /5Alexa647/CCTGTCGCCC*A*C*G*C

Sequence showing the position of primer annealing (underlined sequence). Nucleotides in red bold are the polymorphisms tested for the allele-specific extensions and the ligation reactions.

FWD-982bp

```
1    TCTGCGCTAACCCGCATGCTGCCTGCCCGCCTGCCGCTCACCTGGGACAG

51   AGGACTCGCCGGTGGAGGGGCCTGGCTTCGGGCTCAGTACGGTGTACAGG

101  CGGAGGGCCCTCAGCCGCGTGGCGTGACCAAGTTGGCGGTGGCTGAGGAG

151  TTGGTGGTGGCGGCGTTTTCCTTGCAGCGGCTGGATCCTGCCGTGTGGAC

201  TCTGTGCGGTGCCCGCAGCGGTGCTGGCGCTCGCCTATCGCTCTGCTCTC

251  TCTTTGTAGACGGCGGGCGCTAACACCACCGACAAGGAGCTAGAGGTTCT

301  CTCCTTGCACAACGTCACCTTTGAGGACGCCGGGGAGTACACCTGCCTGG

351  CGGGCAATTCTATTGGGTTTTCTCATCACTCTGCGTGGCTGGTGGTGCTG
```

401 CCAGGTACCGGCTTCTGCTGCTGCTGCTCCGCACTGTCTGGGGGACG

451 CTGGCTCGGGACACGCCAAAGCTGCCAGGACGGACGGGAATCCTGTGACT

FWD 450bp

501 TACGGCCGTCCCGCTTCTTGAGCCCTCACTCCTGGCCCTGTGCCCAGTGT

551 GGGGACAAAGTTGGCCTGGCCCGGTCCTGGTCCCAGAGGGCCCCCTCAGC

601 CCCTCGAGCCCACTTCCATCTGGGTCCCAAAGCCTCTCTGTGGTCTGGTG

651 TCTCCGGGCGCTGGTGGCGGTGTGGGACTGGCTGCTCTCTGGGCTCCTTC

701 TCTCAGGGTCTGGCCTCTAGACTACTGGGGTTACTGACTGCGAGACCCTC

751 CAGACAAGGCGCGTGCAGAGGCTCTGAGCCCCCTTCCGCTCCCAGTGGTG

FWD 175bp

801 CCTGCGGCTCTGGGCCAGGGGCATCCATGGGAGCCCCGTGGGGGGGGGGG

851 CCAGGCCAGGCCTCAACGCCCATGTCTTTTGCAGCCGAGGAGGAGCTGGT

FWD 66bp

SNP1

901 GGAGGCTGACGAGGCGGGCAGTGTGTATGCAGGCATCCTCAGCTAC**G/A**GGG

Reverse SNP1

951 TGGGCTTCTTCCTGTTCATCCTGGTGGTGGCGGCTGTGACGCTCTGCCGC

1001 CTGCGCAGCCCCCCCAAGAAAGGCCTGGGCTCCCCCACCGTGCACAAGAT

1051 CTCCCGCTTCCCGCTCAAGCGACAGGTAACAGAAAGTAGATACCAGGTTC

1101 TGAGCTGCCTGCCCGCCAGGCCTCCTGGAGCCCCACCTCGGCCCACGCTG

1151 GTCCTGGGCTGTGTGAGCCCTCTCTGCAGCCAGGCGGGCTCCCCTCTCCT

1201 CGTCTCTGCTCACCATGTAGAGCCTAGGGTACTTTGGGGCACGAAACATT

1251 CTAAAAATCTTCATTCAATGCTGGTGGAAGTCAGAACGCCCCCCCTTCTG

SNP2

1301 GCCCAGCACTGACCCCCGGCTGTACCTCCACGCCCTGTCGCCCACG**T/C**GGC

Reverse SNP2

1351 GCCAACCTGCCCCTGCTGACC

REFERENCES

(1) Diehl, F.; Li, M.; He, Y.; Kinzler, K. W.; Vogelstein, B.; Dressman, D. *Nat Methods* **2006**, *3*, 551-559.

(2) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. In *Science*, 2005; Vol. 309, pp 1728-1732.

(3) Ritchie, M. E.; Silver, J.; Oshlack, A.; Holmes, M.; Diyagama, D.; Holloway, A.; Smyth, G. K. *Bioinformatics* **2007**, *23*, 2700-2707.

Table S1. Data calculated from ten image frames per dye. The average relative fluorescence is estimated as the intensity gray scale (0-4096) averaged for all beads in the top 20th intensity percentile for each dye in its respective channel after filtering for multi-template beads (a small fraction of the total). The intensity of these bright beads was also measured in the other fluorescent channels since these beads are most likely to exhibit cross-talk between channels. The standard deviation is indicated in parentheses. The average background fluorescence was estimated for each dye in its respective channel as the average intensity of beads with intensities below a threshold of 9 on the log2 scale. This threshold roughly delineates the point in all channels above which the signal derives almost exclusively from the channel of interest, and below which beads have very little PCR product or are empty. The true background would be the autofluorescence of empty beads, which cannot be determined here but is represented by the beads with the lowest signal. The S/N is the signal to noise ratio of the average relative intensity to the average background fluorescence.

Dye	Average relative intensity				Average background intensity	S/N	Exposure time
	A488 channel	A532 channel	A594 channel	A647 channel			
Alexa488	**1388.42 (312.41)**	152.61 (58.98)	233.59 (78.43)	113.86 (22.25)	266.44	5.21	900ms
Alexa532	134.98 (19.41)	**1458.13 (314.00)**	283.19 (52.39)	104.59 (29.78)	206.75	7.05	350ms
Alexa594	139.11 (57.13)	128.18 (61.19)	**1743.79 (412.72)**	147.23 (83.66)	362.69	4.81	80ms
Alexa647	155.35 (32.99)	139.49 (36.39)	315.46 (100.48)	**1066.54 (261.71)**	211.16	5.05	3000ms

Figure S1. A. Density plots and **B.** boxplots of the log2 intensities of beads amplified with various polymerases for different emulsion PCR amplicon lengths.

Figure S1A

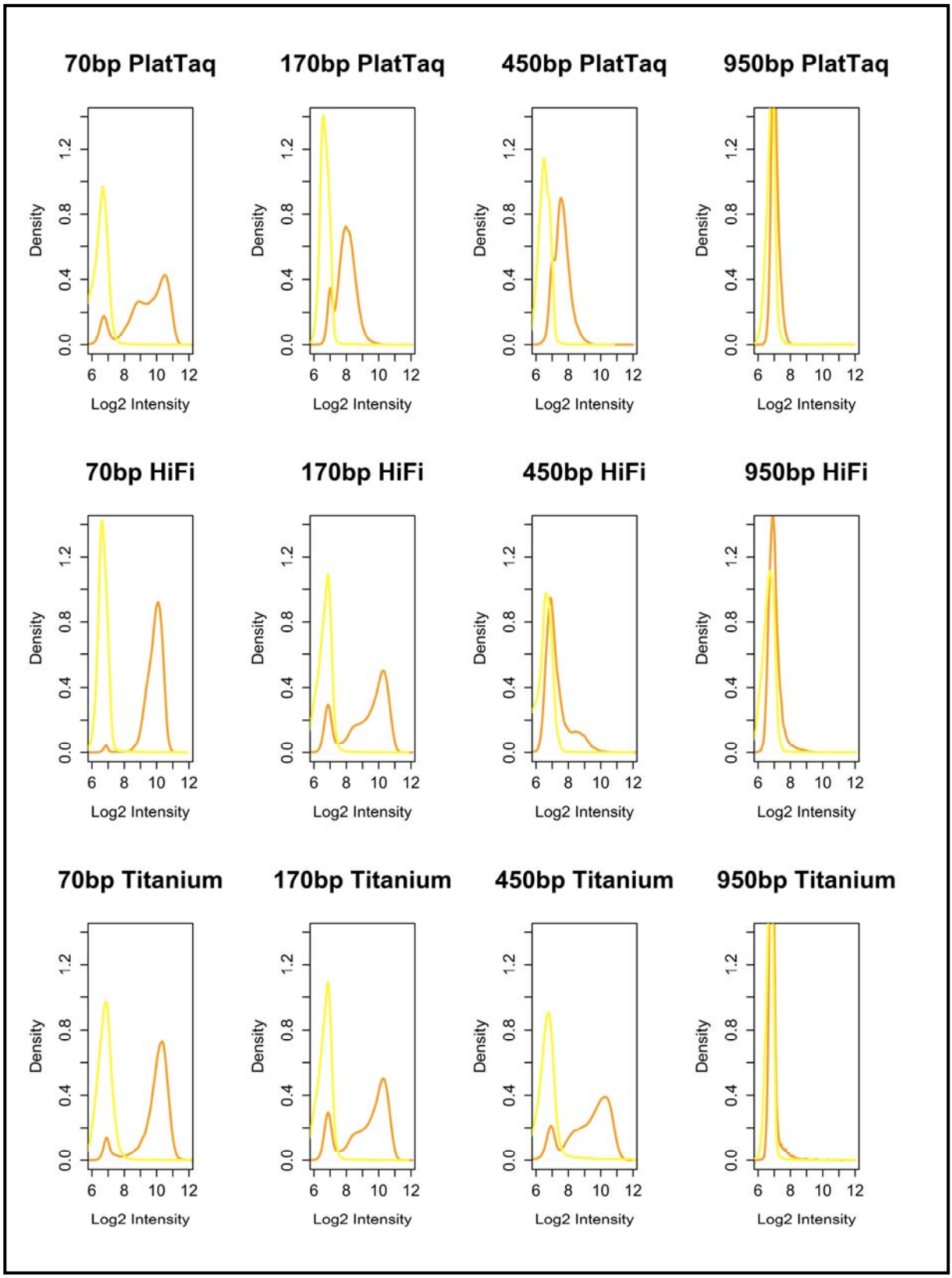


Figure S1B

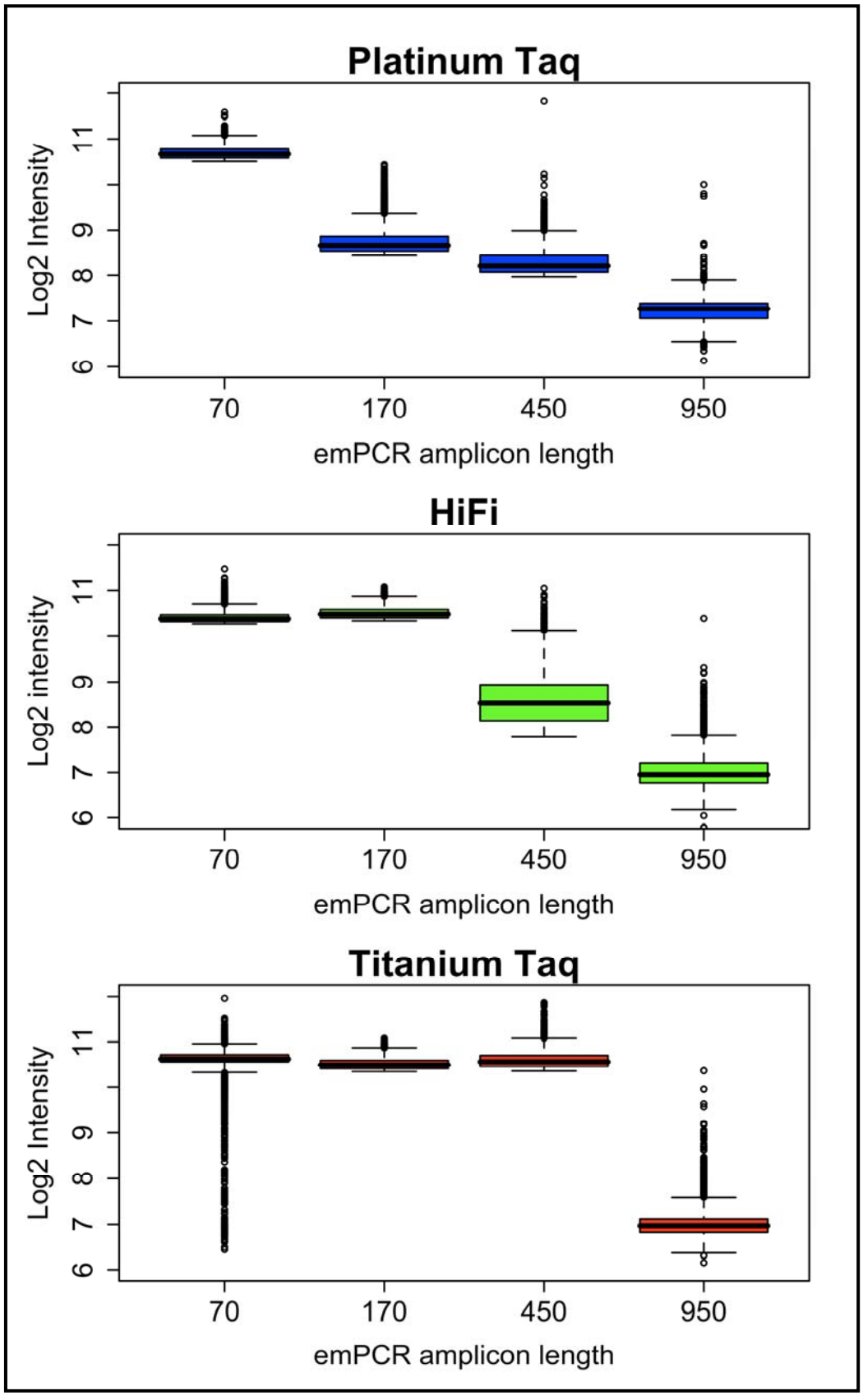
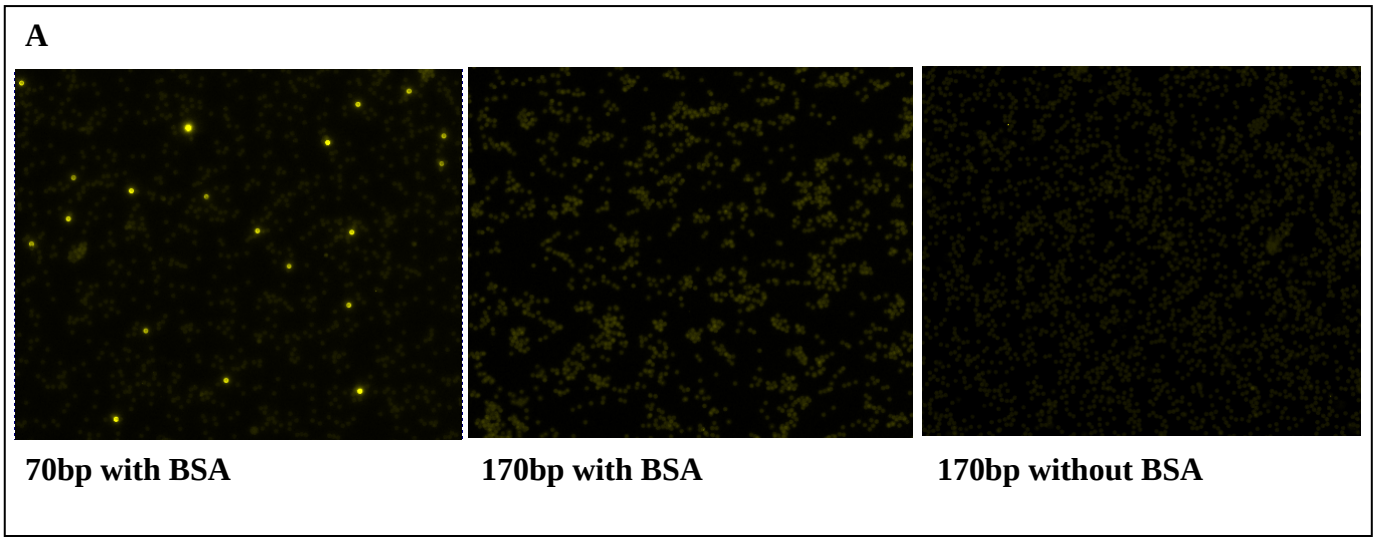

Platinum Taq
Log2 Intensity
6
7
8
9
11
70
170
450
950
emPCR amplicon length
HiFi
Log2 intensity
6
7
8
9
11
70
170
450
950
emPCR amplicon length
Titanium Taq
Log2 Intensity
6
7
8
9
11
70
170
450
950
emPCR amplicon length

Figure S2. Effect of Bovine Serum Albumin (BSA) in the amplification of different sized templates by the Platinum amplification system. **A.** Images of beads obtained with and without the addition 3.75 µg/ml BSA to the amplification of 70bp and 170bp products. The experiment was performed as described for the other product length experiments. **B.** The average intensity of highly intense beads estimated for each experimental condition. Error bars denote one standard deviation from the mean.



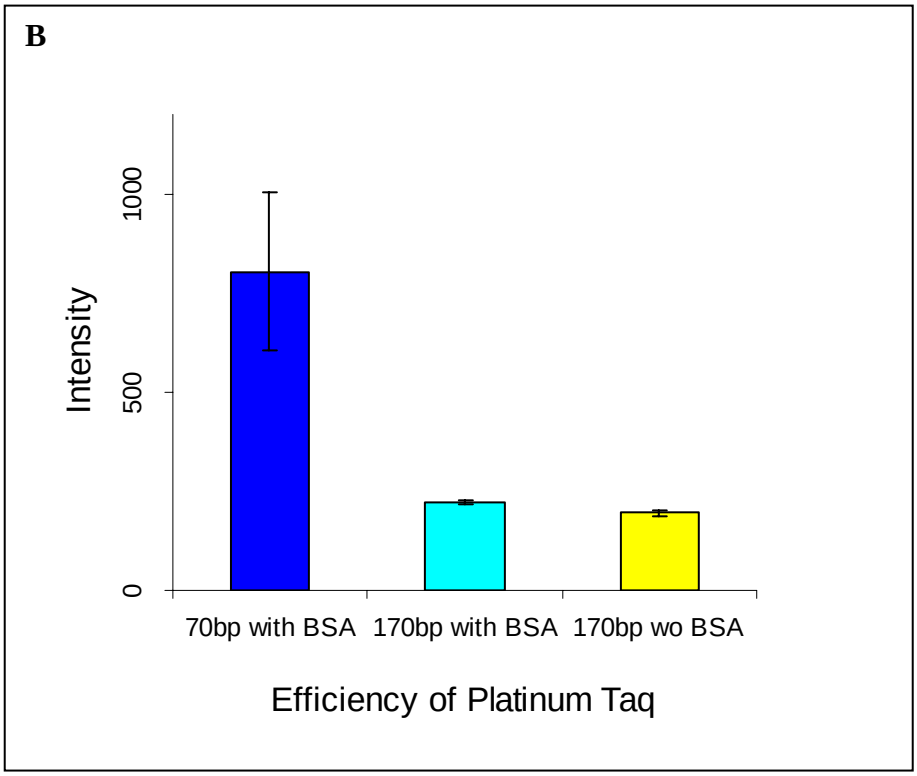


Figure S3. A. Boxplots and **B.** Density plots of the log2 intensities of beads labeled with different dyes imaged in all four channels (channels are labeled as in panel A).

Figure 3A

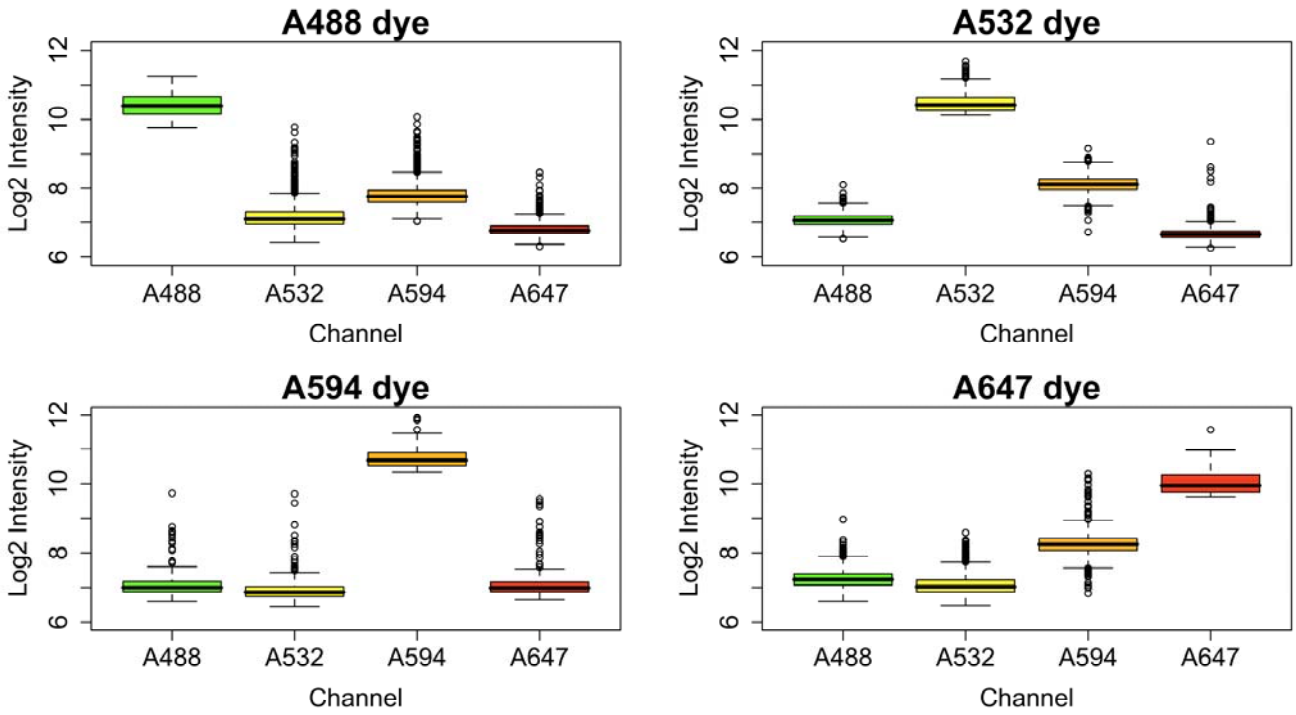


Figure 3B