

**An efficient microarray-based genotyping platform for the identification
of drug-resistance mutations in majority and minority subpopulations
of HIV-1 quasispecies**

S1 File: Supporting information

Verónica Martín^{1,#a}, Celia Perales^{1,2,3}, María Fernández-Algar⁴, Helena G. Dos Santos^{1,#b}, Patricia Garrido^{5,#c}, María Pernas^{5,#d}, Víctor Parro⁴, Miguel Moreno⁴, Javier García-Pérez⁶, José Alcamí⁶, José Luis Torán^{5,#e}, David Abia¹, Esteban Domingo^{1,2}, Carlos Briones^{2,4,*}.

¹ Centro de Biología Molecular ‘Severo Ochoa’ (CBMSO, CSIC-UAM). Campus de Cantoblanco, 28049 Madrid, Spain.

² Centro de Investigación Biomédica en Red de enfermedades hepáticas y digestivas (CIBERehd), Spain.

³ Liver Unit, Internal Medicine, Laboratory of Malalties Hepàtiques, Vall d’Hebron Institut de Recerca-Hospital Universitari Vall d’Hebron (VHIR-HUVH), Universitat Autònoma de Barcelona. 08035 Barcelona, Spain.

⁴ Department of Molecular Evolution, Centro de Astrobiología (CAB, CSIC-INTA). 28850 Torrejón de Ardoz, Madrid, Spain.

⁵ Biotherapix, SLU. Parque Tecnológico de Madrid, 28760 Tres Cantos, Madrid. Spain.

⁶ AIDS Immunopathogenesis Unit, Instituto de Salud Carlos III. 28220 Majadahonda, Madrid, Spain.

^{#a}Current Address: Centro de Investigación en Sanidad Animal (CISA), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). 28130 Valdeolmos, Madrid, Spain.

^{#b}Current Address: Department of Biological Sciences, Florida International University. Miami, FL 33199, USA.

^{#c}Current Address: Sygnis Biotech. Parque Científico de Madrid. Campus de Cantoblanco, 28049 Madrid, Spain.

^{#d}Current Address: Instituto de Salud Carlos III. 28220 Majadahonda, Madrid, Spain.

^{#e}Current Address: Spanish National Center for Biotechnology (CNB, CSIC). Campus de Cantoblanco, 28049 Madrid, Spain.

* Corresponding author

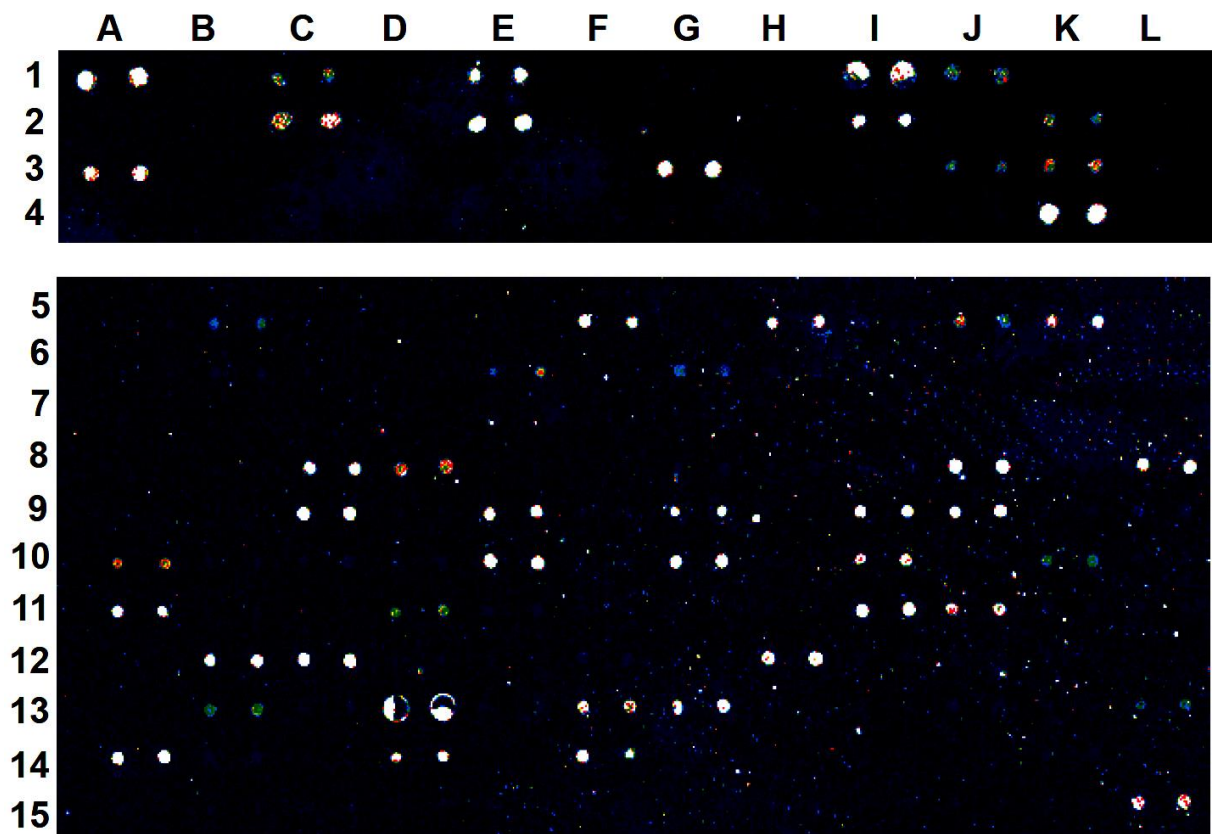
E-mail: cbriones@cab.inta-csic.es

Figure A. Scheme of the HIV-1 genotyping microarray and examples of the hybridization signals produced by wild type target DNA molecules. (a) Position of the 360 spots printed on each of the four subarrays of the microarray. Discriminating probes are shown in black, positive HIV-1 internal hybridization controls (IHC-PR3, IHC-PR4, IHC-RT3 and IHC-RT4) are depicted in green, negative controls (C-:FMDV-G142-15r and C-:FMDV-E142-15r, corresponding to sequences of the foot-and-mouth disease virus VP1 coding region) appear in red and additional negative controls including spotting solution (C-) are shown in grey. Yellow background highlights the probes complementary to wild type sequences for each queried codon. The highly performing probes that were maintained after the quality control protocol (see main text) are underlined. Further details on the probe sequences are given in Supp. Table S1. Rows 1-4 contain the PR-specific probes and rows 5-15 the RT ones. (b) and (c) Examples of microarray images corresponding to the hybridization of PR and RT targets with wild type sequences, using ScanArray and GenePix microarray scanners (see Methods), respectively.

a)

	A	B	C	D	E	F	G	H	I	J	K	L
1	IHC-PR3 • •	IHC-PR4 • •	<u>R8-PR-2</u> • •	<u>Q8-PR-2</u> • •	<u>D30-PR</u> • •	N30-PR • •	<u>M36-PR</u> • •	I36-PR • •	<u>M46b-PR</u> • •	<u>M46c-PR</u> • •	<u>F46-PR</u> • •	<u>I46-PR</u> • •
2	<u>L46-PR</u> • •	<u>V46-PR</u> • •	<u>G48-PR</u> • •	<u>V48-PR</u> • •	<u>I50-PR</u> • •	L50a-PR • •	<u>L50b-PR</u> • •	<u>V50-PR-2</u> • •	<u>I54-PR</u> • •	<u>V54-PR-2</u> • •	<u>A71-PR-2</u> • •	<u>V71-PR-2</u> • •
3	<u>V82b-PR-2</u> • •	<u>A82-PR-2</u> • •	<u>F82-PR-2</u> • •	<u>I82-PR-2</u> • •	<u>S82-PR-2</u> • •	<u>T82-PR-2</u> • •	<u>I84-PR-2</u> • •	<u>A84-PR-2</u> • •	<u>V84-PR-2</u> • •	<u>L90a-PR</u> • •	<u>L90a-PR-2</u> • •	<u>I90-PR</u> • •
4	I90-PR-2 • •	M90-PR • •	<u>M90-PR-2</u> • •	C-: FMDV- G142-15r • •	C-: FMDV- E142-15r • •	C- • •	C- • •	C- • •	C- • •	C- • •	IHC-PR3 • •	IHC-PR4 • •
5	IHC-RT3 • •	IHC-RT4 • •	<u>M41</u> • •	<u>L41a</u> • •	L41b • •	<u>A62</u> • •	<u>V62</u> • •	<u>K65-2</u> • •	<u>R65-3</u> • •	<u>D67a-2</u> • •	<u>D67b-2</u> • •	<u>E67-2</u> • •
6	<u>G67</u> • •	<u>G67-2</u> • •	<u>N67-2</u> • •	<u>Del67</u> • •	<u>S68</u> • •	<u>N68</u> • •	<u>T69b</u> • •	<u>T69c</u> • •	<u>A69</u> • •	<u>D69</u> • •	<u>G69</u> • •	<u>S69a</u> • •
7	<u>S69b</u> • •	<u>S69R70</u> • •	Ins69a • •	<u>Ins69b</u> • •	<u>Ins69c</u> • •	<u>Ins69d</u> • •	<u>Ins69e</u> • •	<u>Ins69f</u> • •	<u>Ins69g</u> • •	<u>Ins69h</u> • •	<u>Ins69i</u> • •	<u>Ins69j</u> • •
8	<u>Ins69k</u> • •	<u>Ins69l</u> • •	<u>K70a</u> • •	<u>K70b</u> • •	<u>E70</u> • •	N70a • •	<u>N70b</u> • •	R70a • •	<u>R70b</u> • •	<u>L74</u> • •	<u>V74-2</u> • •	<u>V75</u> • •
9	<u>I75</u> • •	<u>T75</u> • •	<u>F77</u> • •	<u>L77</u> • •	<u>L100-2</u> • •	<u>I100-2</u> • •	<u>K101-3</u> • •	<u>E101-3</u> • •	<u>K103a-2</u> • •	<u>K103c-2</u> • •	<u>N103-3</u> • •	<u>R103-3</u> • •
10	<u>V106-2</u> • •	<u>A106-2</u> • •	<u>I106-2</u> • •	<u>L106-2</u> • •	<u>V108-2</u> • •	<u>I108-2</u> • •	<u>F116</u> • •	<u>Y116-3</u> • •	<u>Q151-2</u> • •	<u>M151-2</u> • •	<u>I178</u> • •	<u>M178</u> • •
11	<u>V179-2</u> • •	<u>D179-2</u> • •	<u>E179-2</u> • •	<u>Y181</u> • •	<u>C181</u> • •	<u>H181</u> • •	<u>I181</u> • •	<u>L181</u> • •	<u>M184a</u> • •	<u>M184b</u> • •	<u>I184</u> • •	<u>T184</u> • •
12	<u>V184b</u> • •	<u>Y188a</u> • •	<u>Y188c</u> • •	<u>C188-2</u> • •	<u>H188</u> • •	L188a • •	<u>L188b</u> • •	<u>G190</u> • •	<u>A190</u> • •	<u>E190</u> • •	<u>Q190</u> • •	<u>S190</u> • •
13	<u>T190</u> • •	<u>L210-3</u> • •	<u>W210-2</u> • •	<u>R211-2</u> • •	<u>K211-2</u> • •	<u>T215a-2</u> • •	<u>T215b-2</u> • •	<u>C215-2</u> • •	<u>F215-2</u> • •	<u>S215-2</u> • •	<u>Y215-2</u> • •	<u>K219</u> • •
14	<u>K219-2</u> • •	<u>E219</u> • •	<u>Q219</u> • •	<u>P225-2</u> • •	<u>H225-2</u> • •	<u>M230-3</u> • •	<u>L230-3</u> • •	<u>P236-2</u> • •	<u>L236-2</u> • •	<u>K238</u> • •	<u>T238</u> • •	C-: FMDV- G142-15r • •
15	C-: FMDV- E142-15r • •	C- • •	C- • •	C- • •	C- • •	C- • •	C- • •	C- • •	C- • •	C- • •	IHC-RT3 • •	IHC-RT4 • •

b)



c)

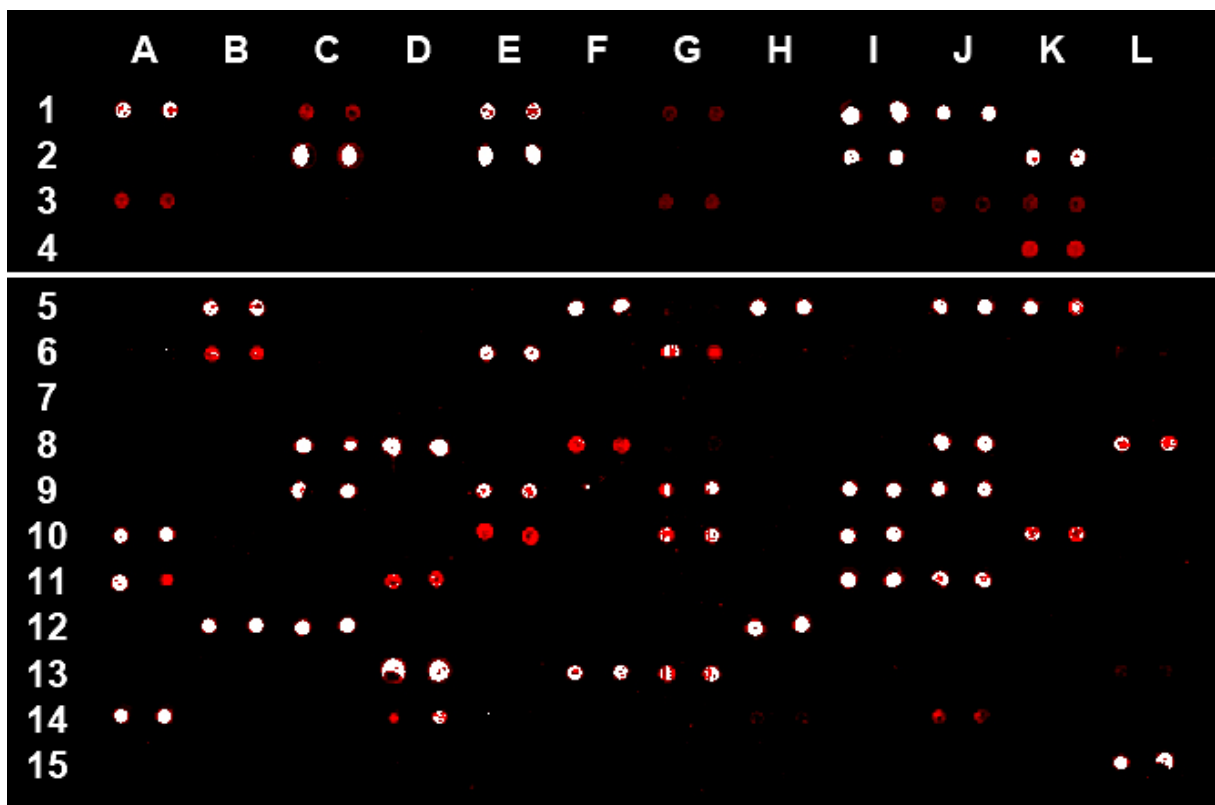
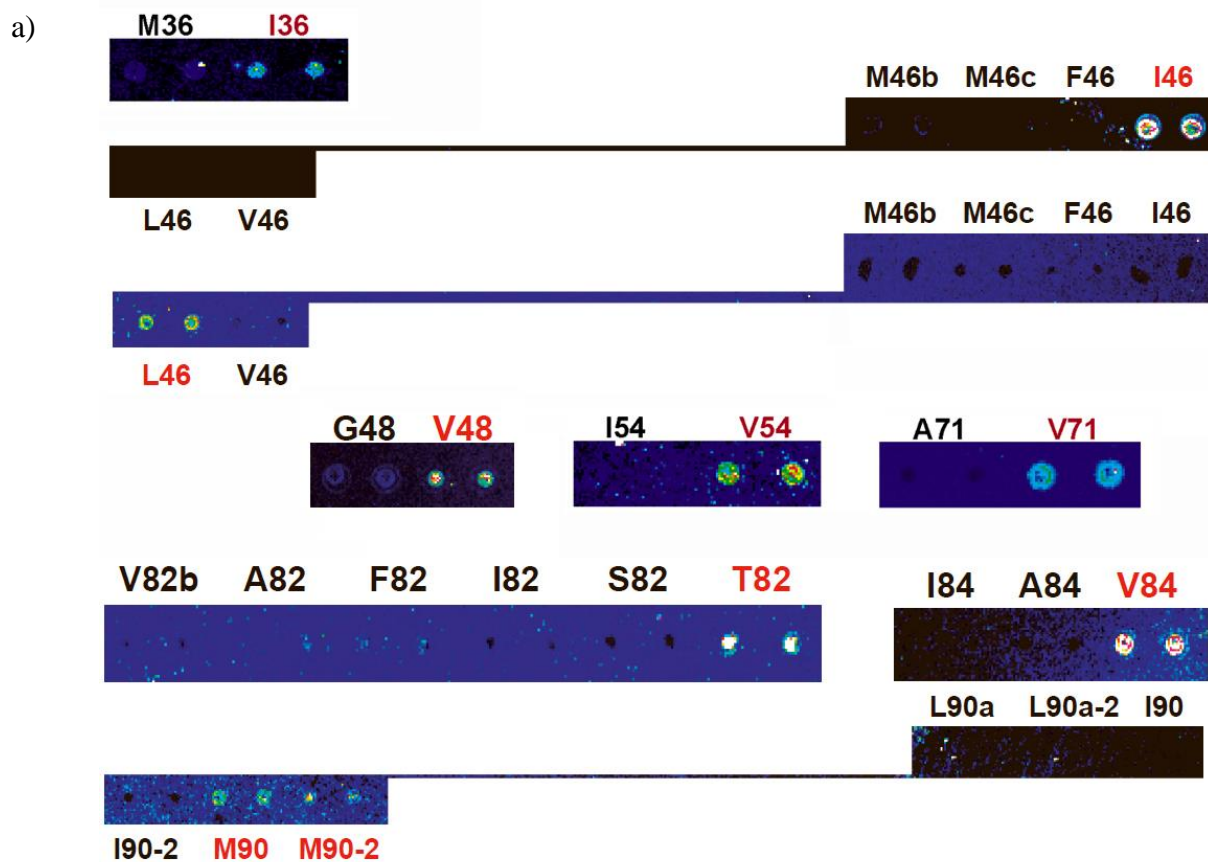


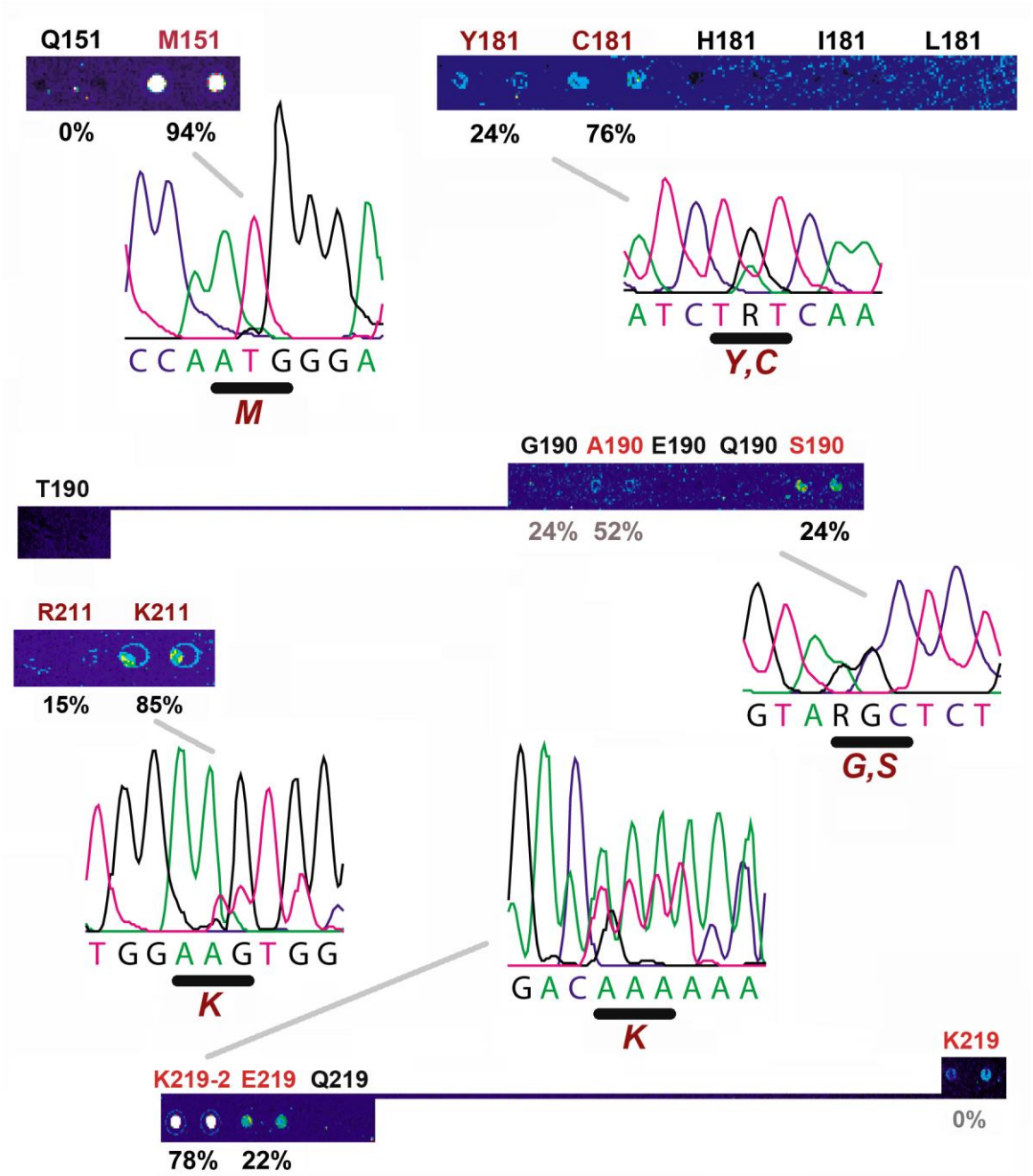
Figure B. Examples of the hybridization results of different PR and RT targets including drug resistance mutations and insertions. (a) Discrimination of mutants in the HIV-1 PR: I36, I46, L46, V48, V54, V71, T82, V84, M90+M90-2 (highlighted in red). (b) Discrimination of mutant and insert-containing targets of the HIV-1 RT: V62, S68, S69, Ins69a, Ins69b, Ins69c, Ins69d, I75, T75, I100, E101, I108, M151, M178, C181, I184, H188, K211, Y215, C215 and T238. The whole set of raw microarray hybridization data used in this study has been uploaded to Gene Expression Omnibus (GEO) repository (<https://www.ncbi.nlm.nih.gov/geo/>), with the accession number GSE90621.



b)



Figure C. Examples of the hybridization results produced by different RT targets belonging to the clinical samples of the test set that contained majority and minority variants. Positive signals in the microarray are highlighted in red. The theoretical percentage of signals at each probe position, derived from the analysis of clonal sequences (13 to 34 clones per sample) is shown below the hybridization signals. Different situations have been chosen to exemplify either concordant or discrepant results between the hybridization signals and their expected percentage derived from clonal analysis (discrepant percentages are depicted in grey). The sequencing electropherograms corresponding to the population sequence are also shown, together with the derived amino acid(s) assumed to be present at the interrogating codon.



b)

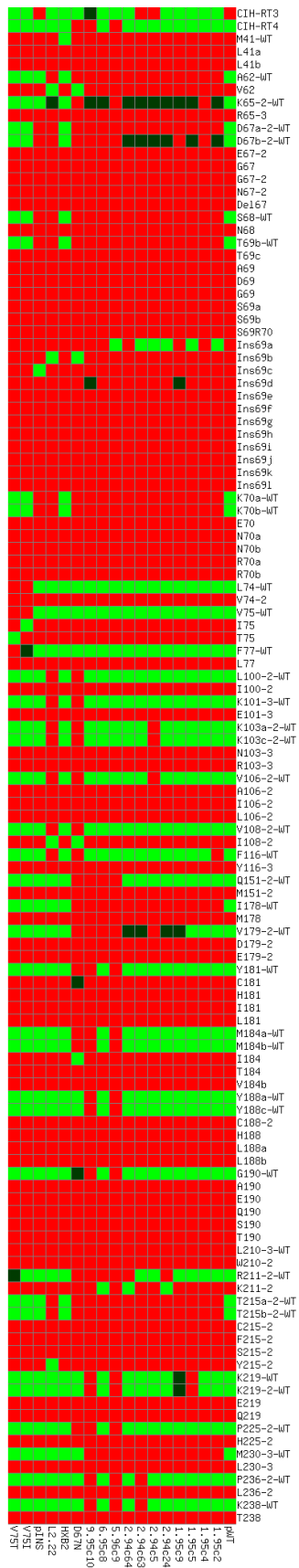


Figure E. Density of the raw hybridization data for three key variables quantified by *GenePix* and *ScanArray* scanners. The threshold values are shown with dotted lines. Color code: Green, true positive (TP) signal; Red, false positive (FP); Blue, true negative (TN); Black, false negative (FN).

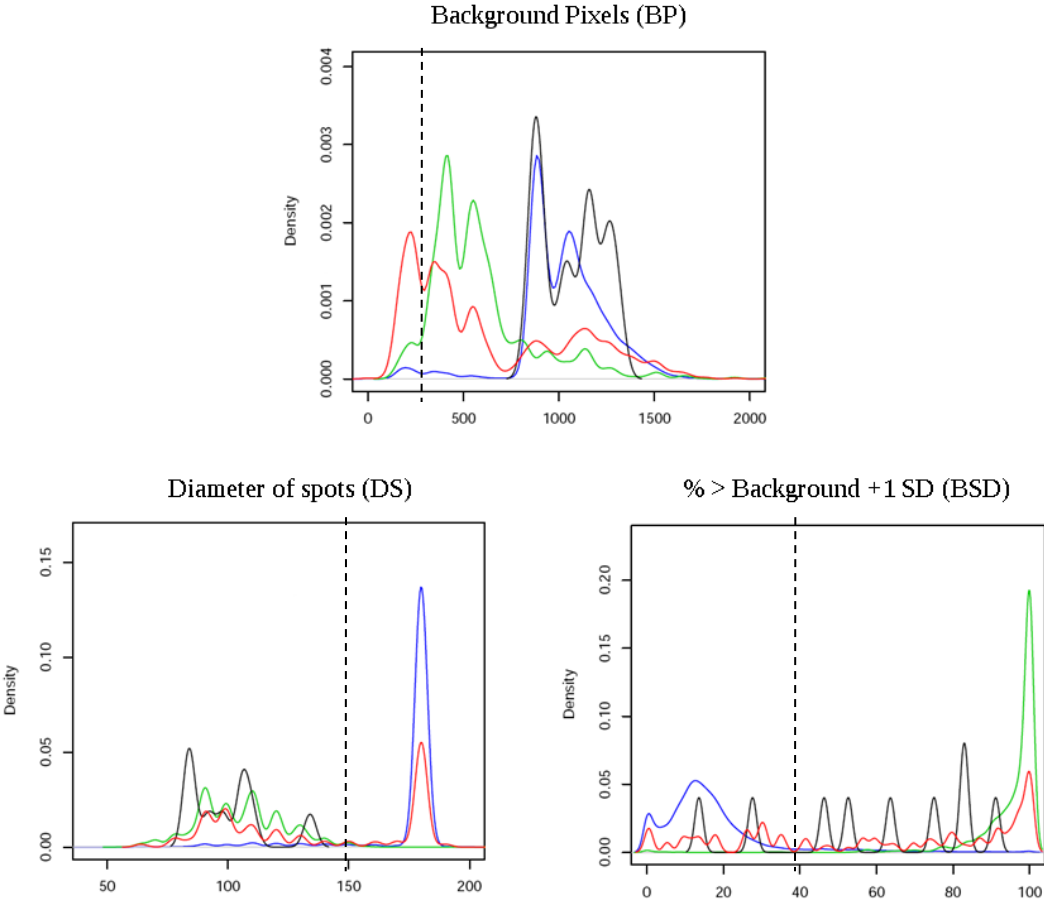
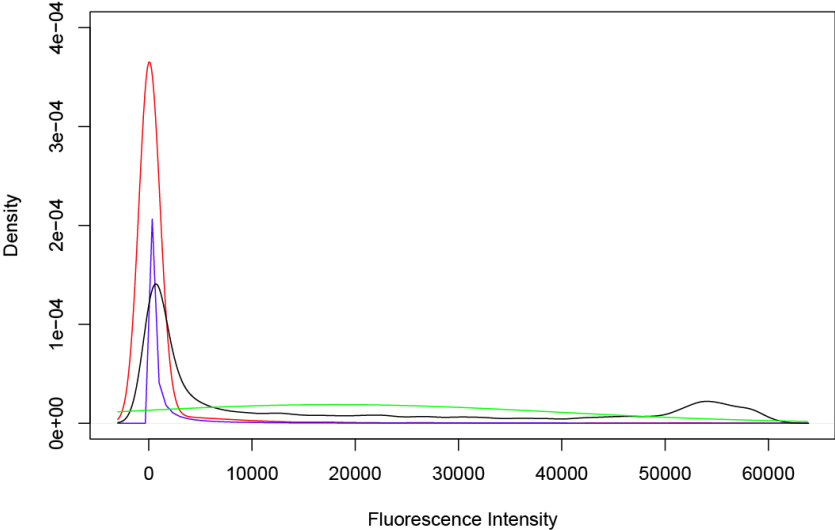


Figure F. Test of the normalization performance of data from the training set. a) Density of positive and negative raw fluorescence data. b) Normalized fluorescence data using the mean positive signal of each array region (RT or RT): distributions of positive signals are centred in value 1 of normalized intensity, while distributions of negative signals are close to zero. Color code: Red, density of negative hybridization data; Blue, fit to a log-normal distribution of negative data; Black, density of positive hybridization data; Green, fit to a normal distribution of positive data.

a)



b)

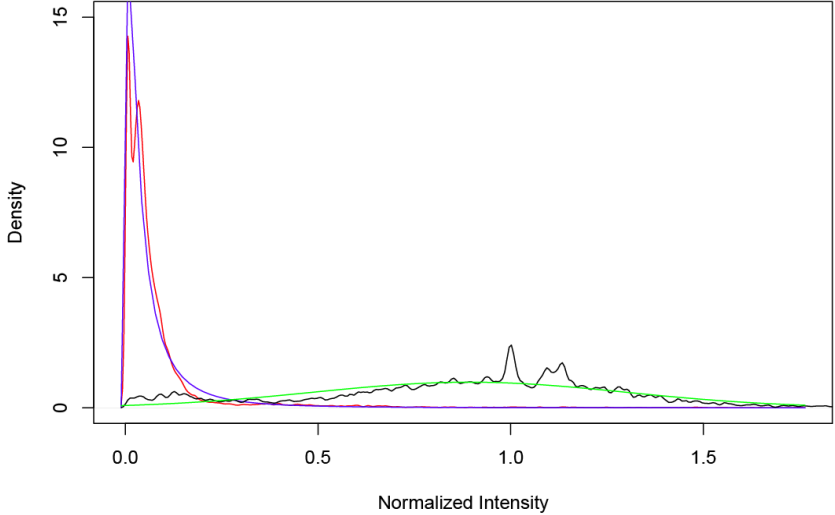
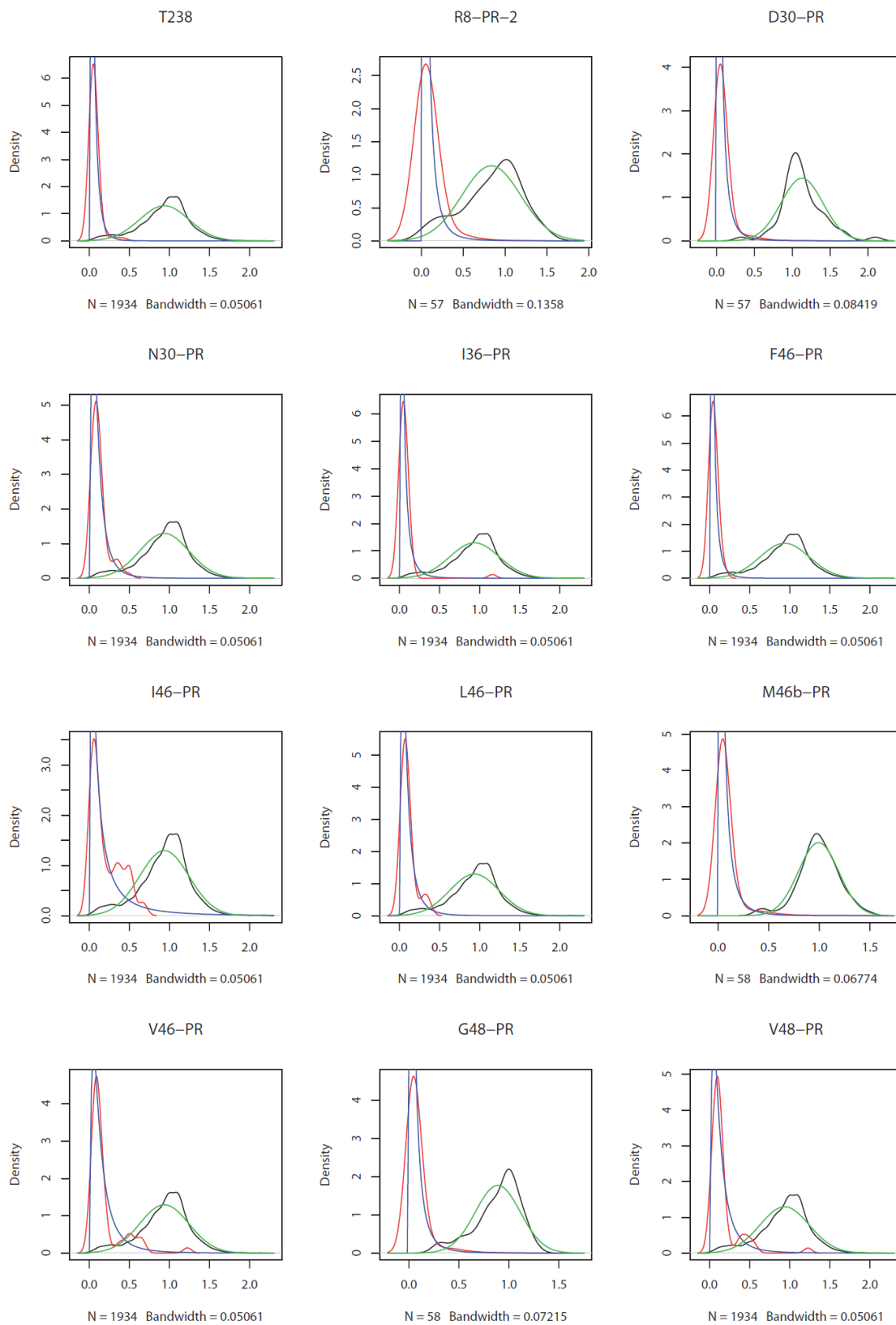
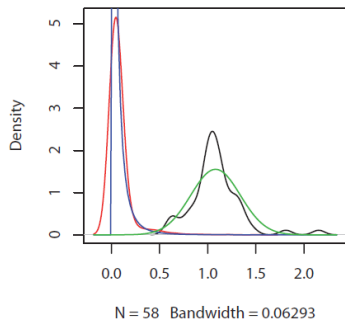


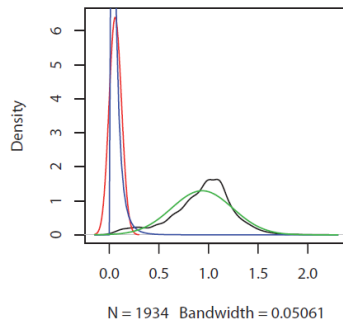
Figure G. Probe-specific curves for the 124 spotted probes that passed the quality control. Density of normalized hybridization signals from the training set with their fitted distribution functions for positive and negative data. Color code: see Supp. Fig S6.



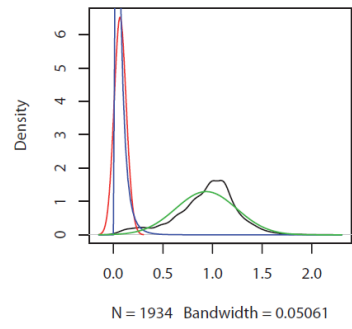
I50-PR



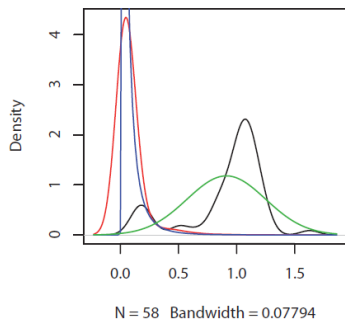
L50b-PR



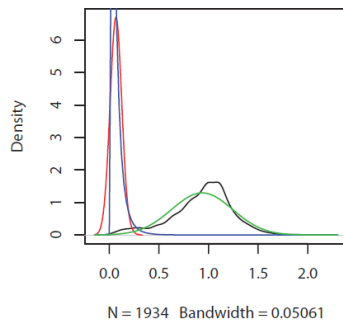
V50-PR-2



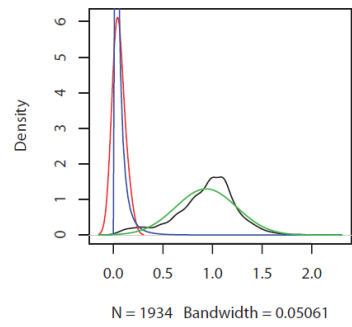
I54-PR



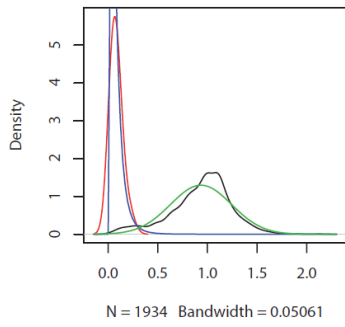
V54-PR-2



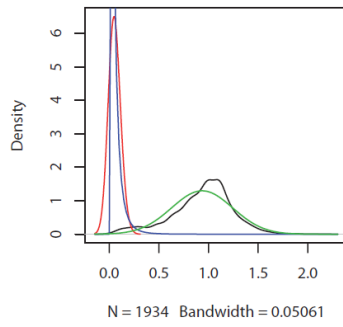
V71-PR-2



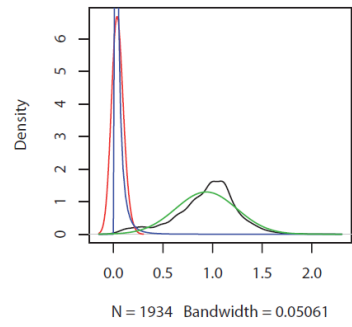
A82-PR-2



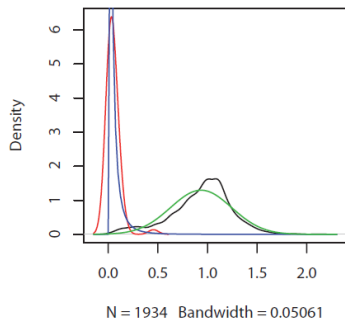
F82-PR-2



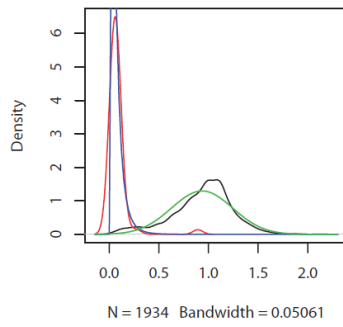
I82-PR-2



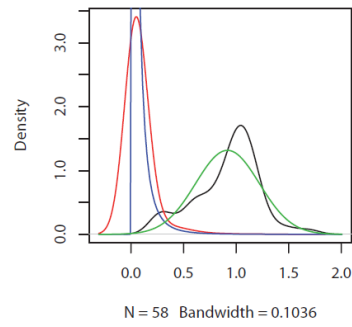
S82-PR-2



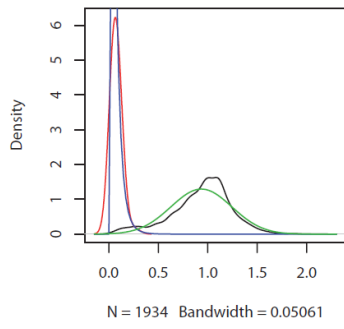
T82-PR-2



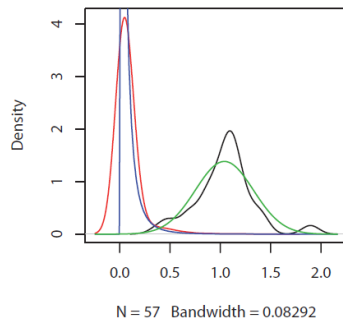
V82b-PR-2



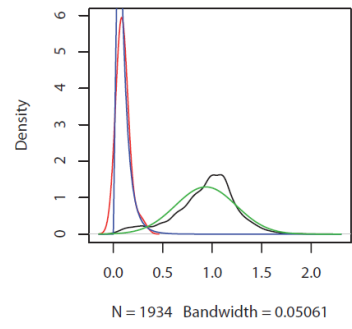
A84-PR-2



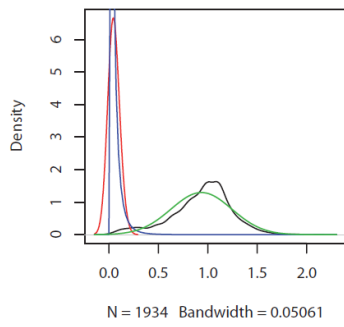
I84-PR-2



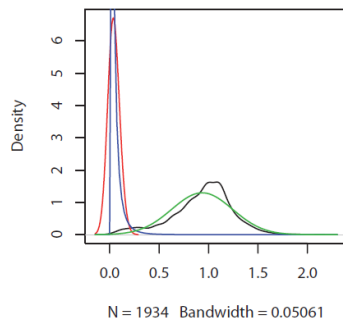
V84-PR-2



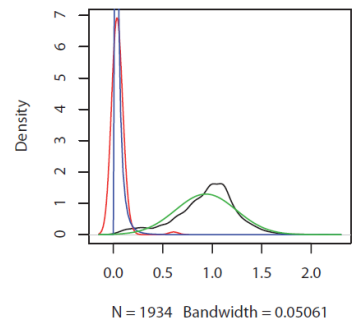
I90-PR



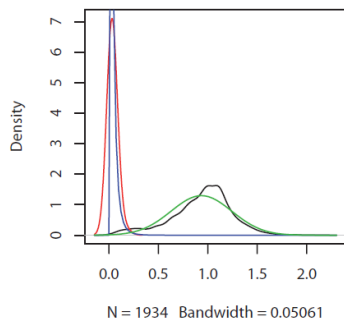
M90-PR-2



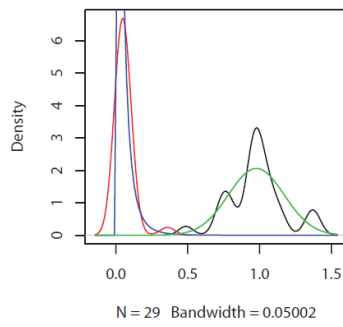
L41a



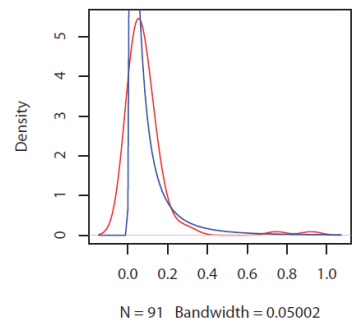
M41



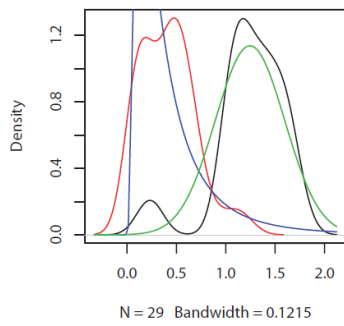
A62



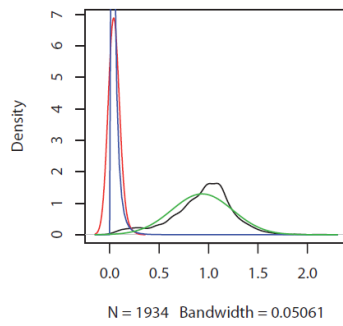
V62



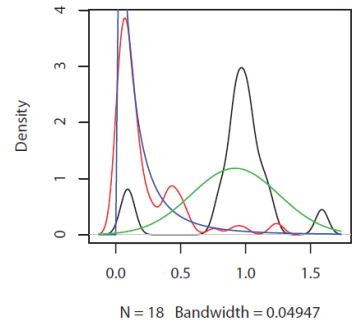
K65-2



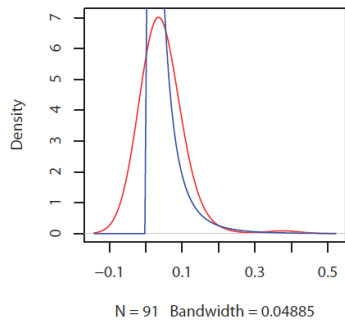
R65-3



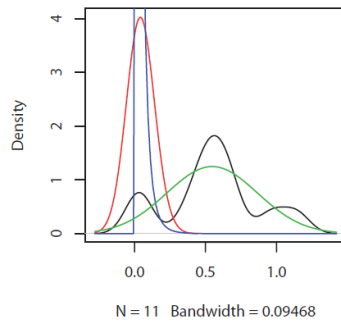
D67a-2



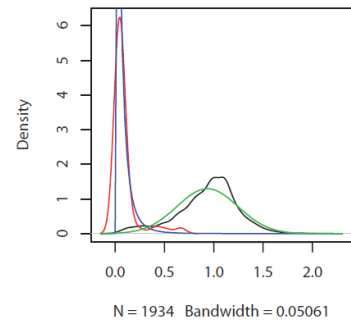
Ins69b



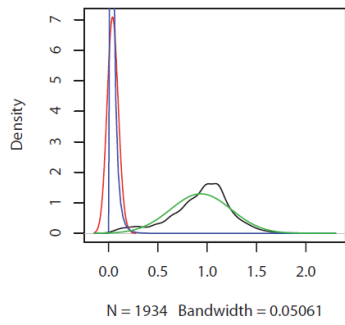
Ins69c



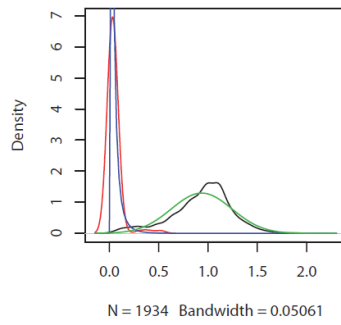
Ins69d



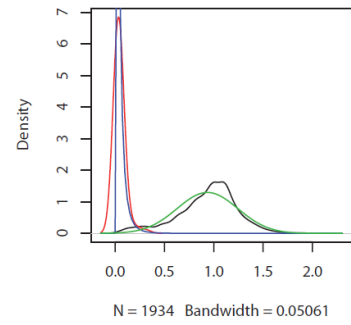
Ins69e



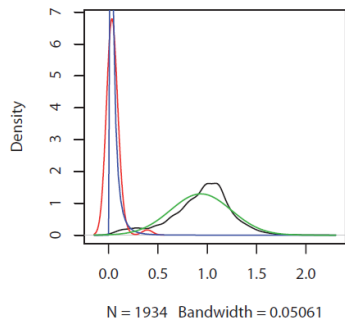
Ins69f



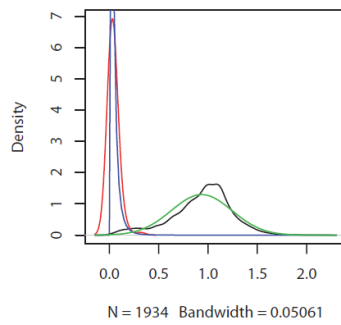
Ins69g



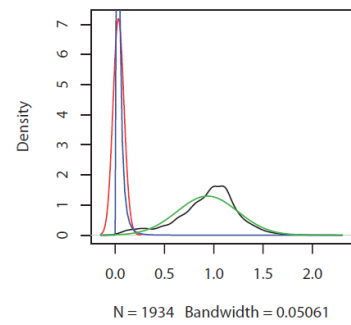
Ins69h



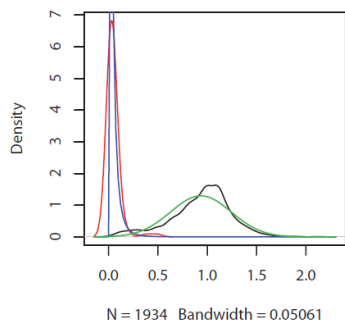
Ins69i



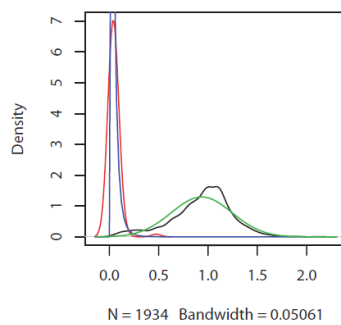
Ins69j



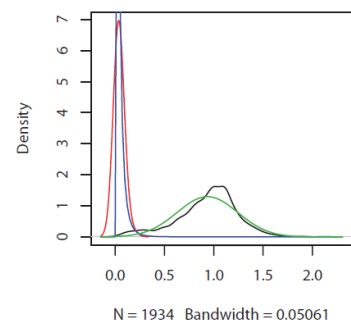
Ins69k



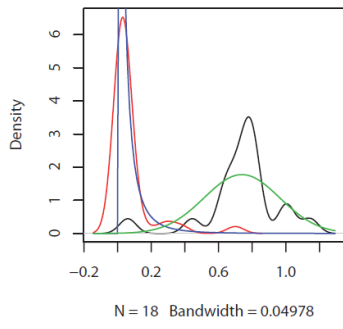
Ins69l



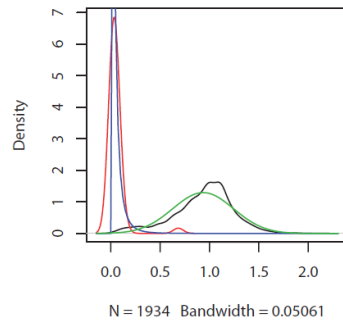
E70



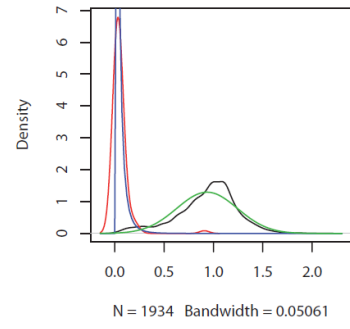
K70a



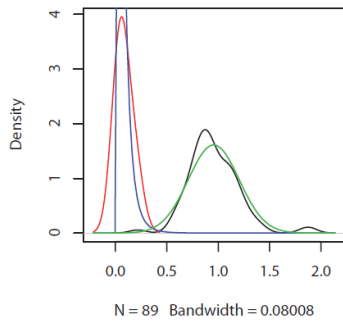
N70b



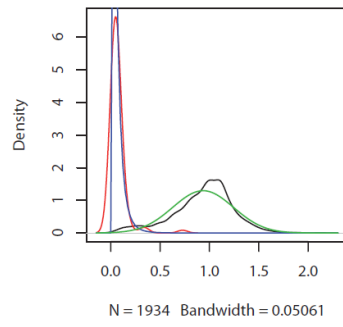
R70b



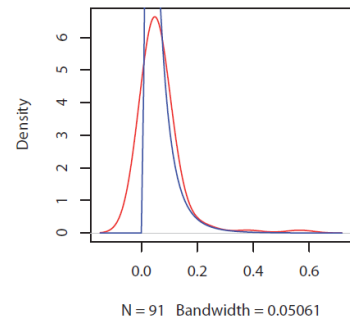
L74



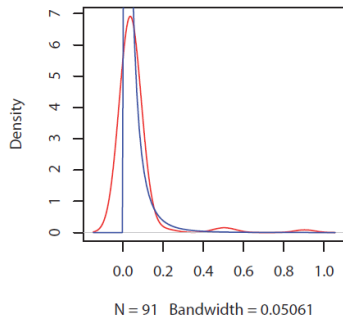
V74-2



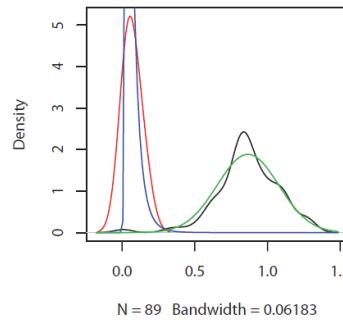
I75



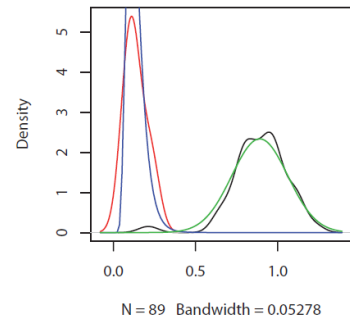
T75



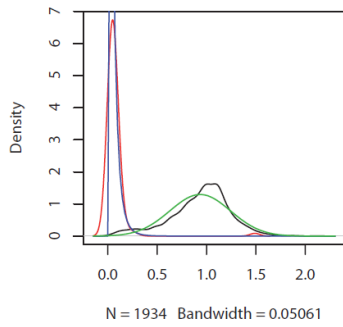
V75



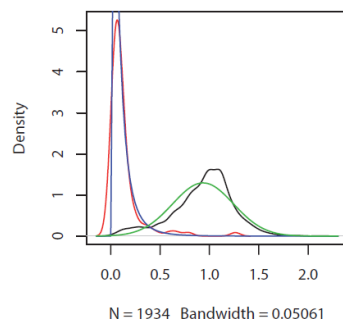
F77



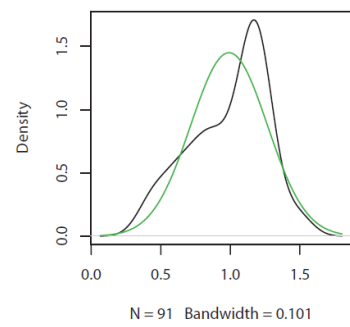
L77



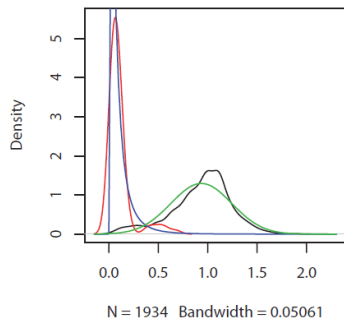
I100-2



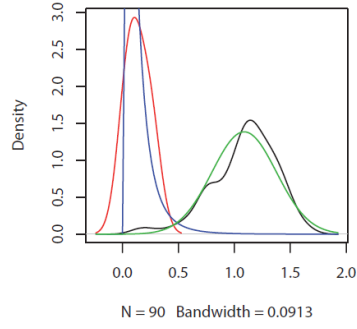
L100-2



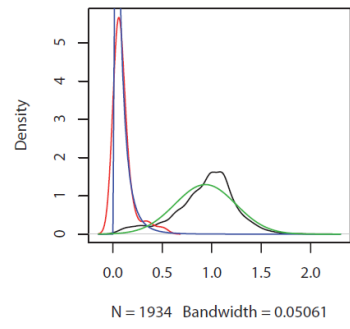
E101-3



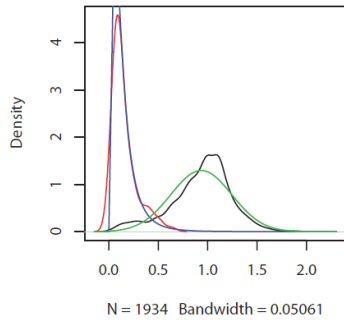
K103a-2



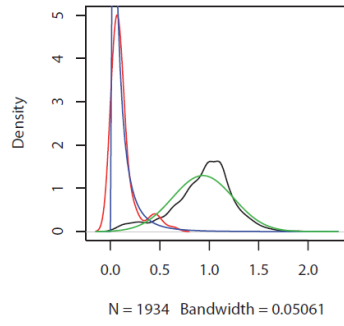
N103-3



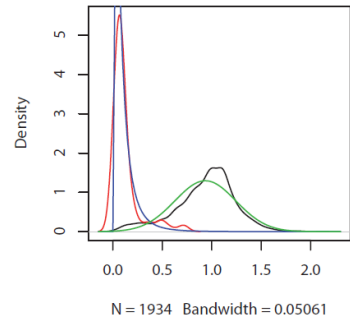
R103-3



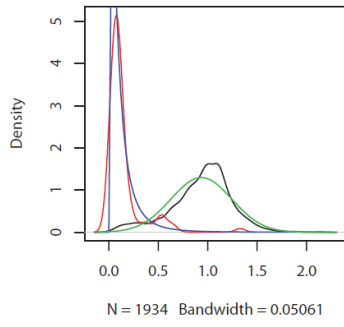
A106-2



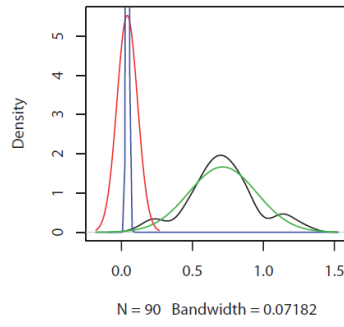
I106-2



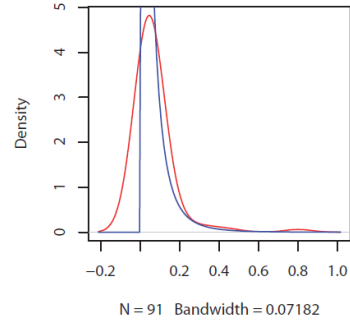
L106-2



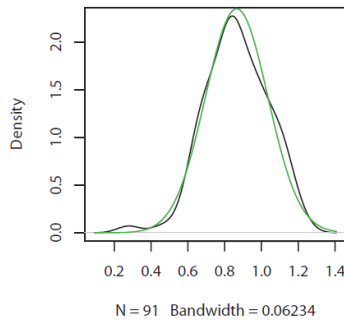
V106-2



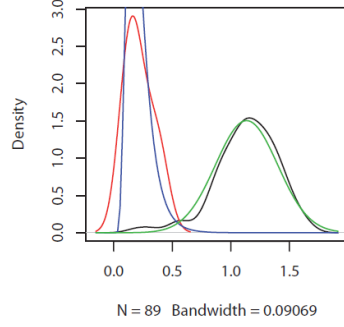
I108-2



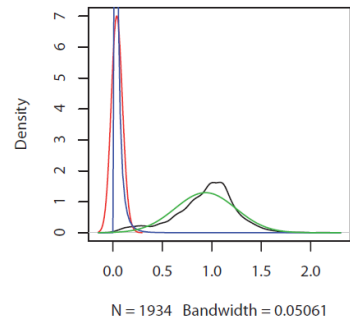
V108-2



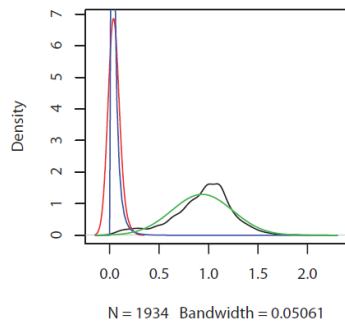
F116



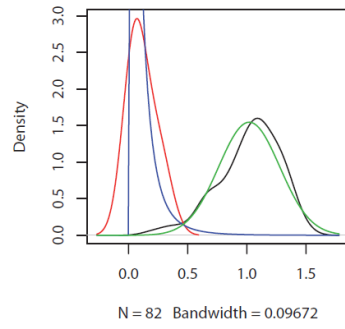
Y116-3



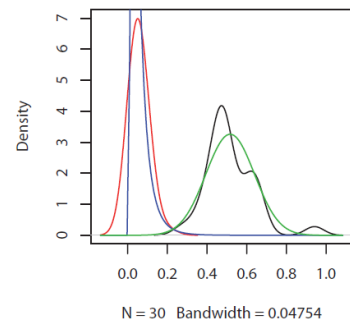
M151-2



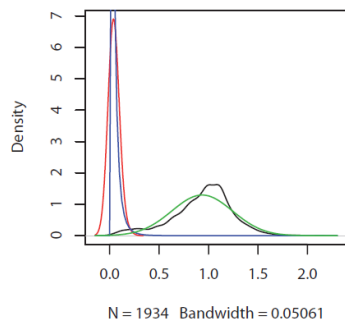
Q151-2



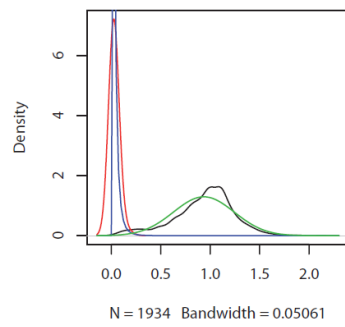
I178



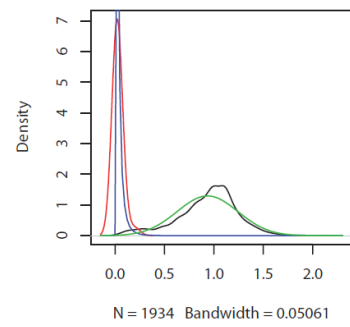
M178



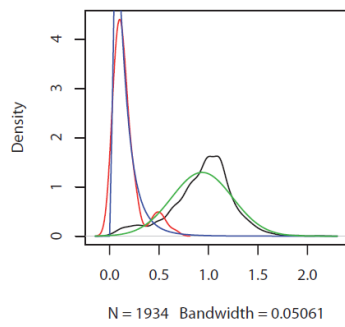
D179-2



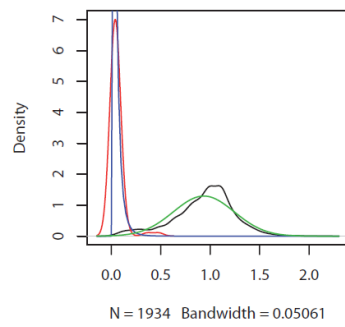
E179-2



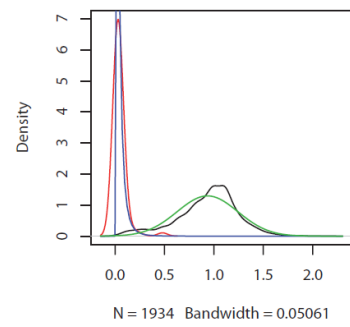
C181



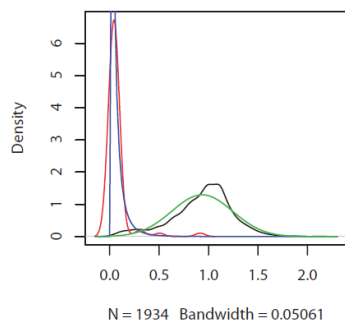
H181



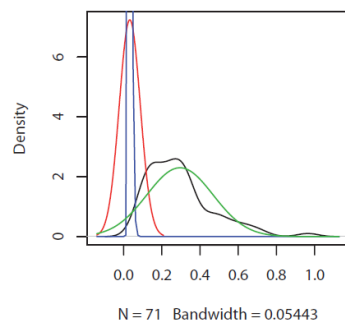
I181



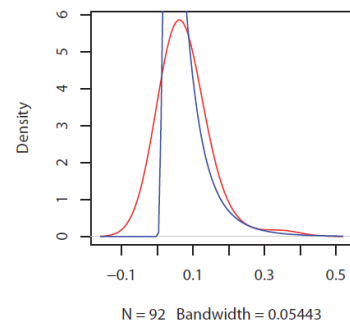
L181



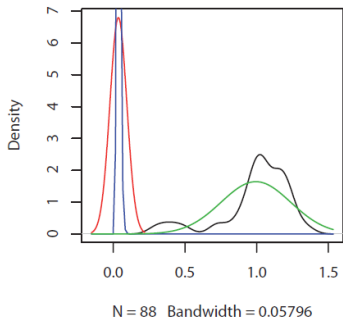
Y181



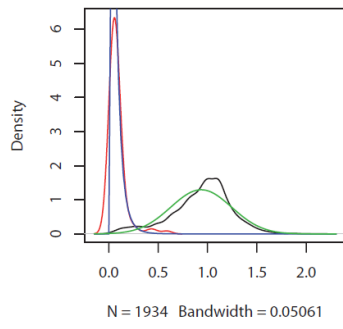
I184



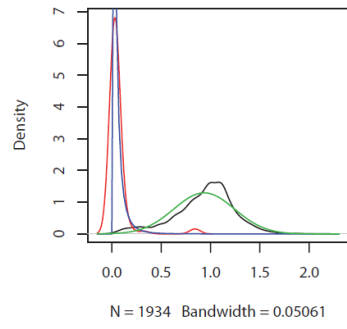
M184a



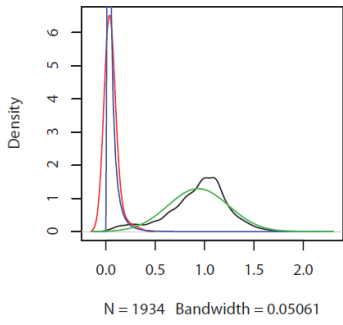
T184



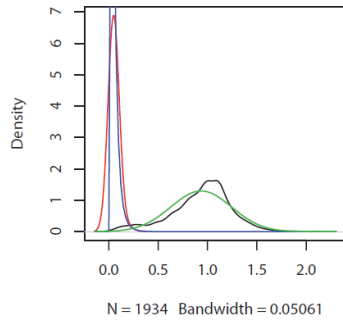
V184b



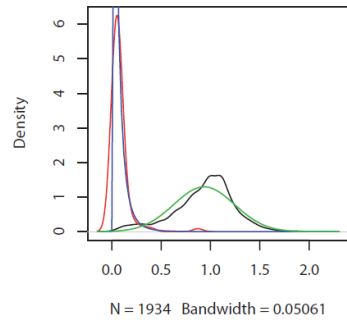
C188-2



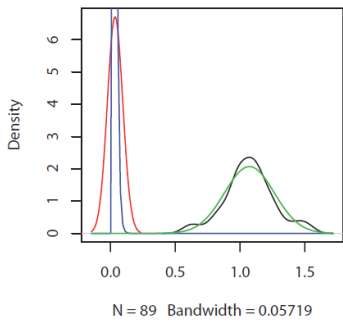
H188



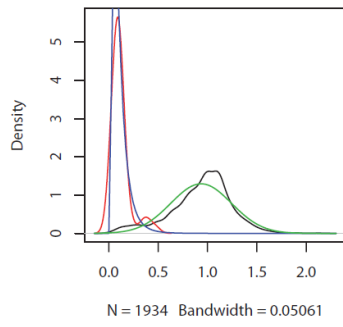
L188b



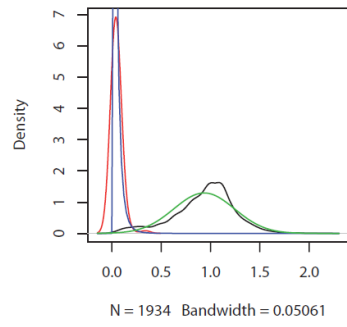
Y188a



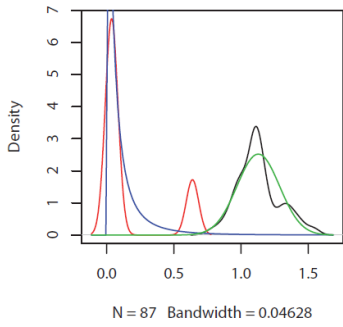
A190



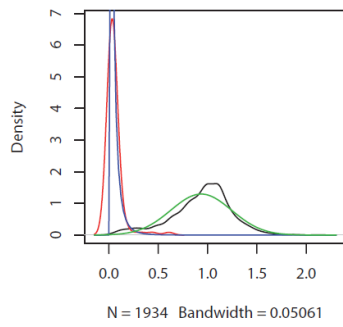
E190



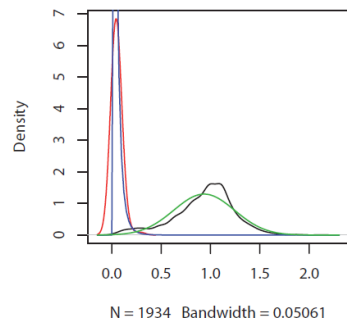
G190



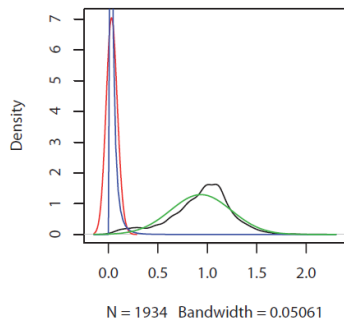
Q190



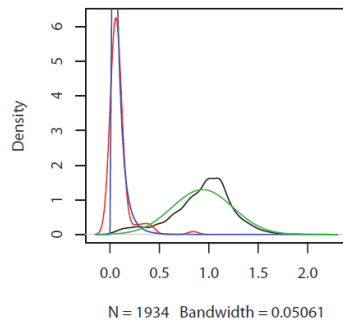
S190



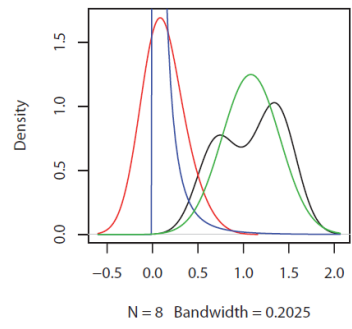
T190



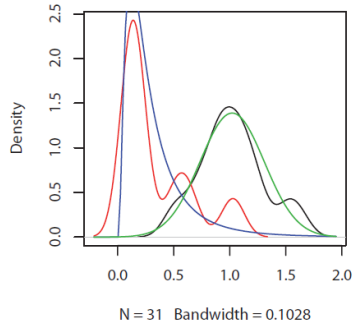
W210-2



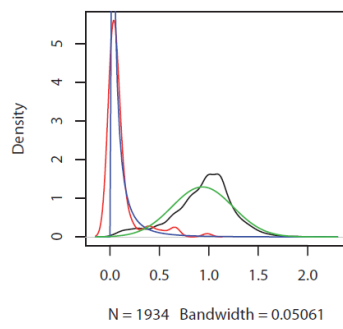
K211-2



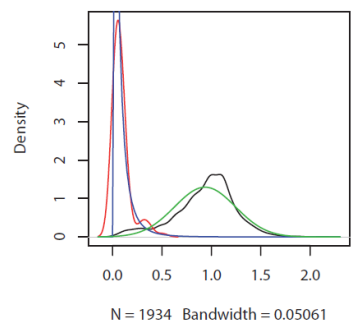
R211-2



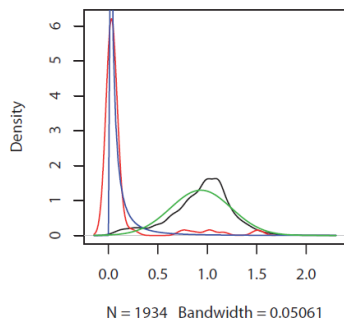
C215-2



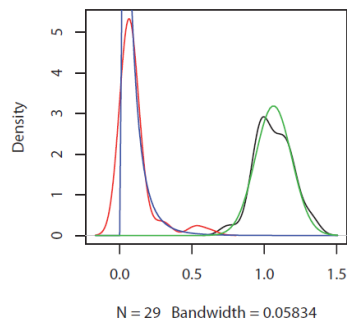
F215-2



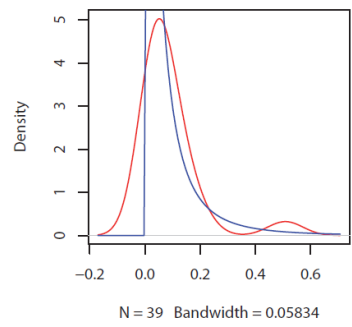
S215-2



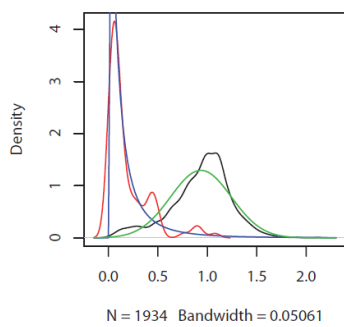
T215b-2



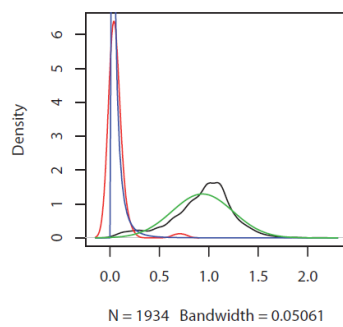
Y215-2



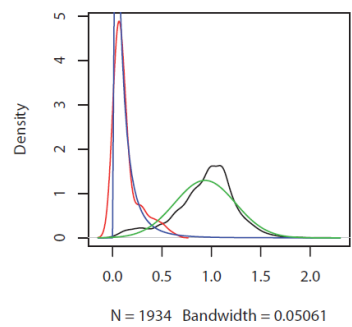
E219



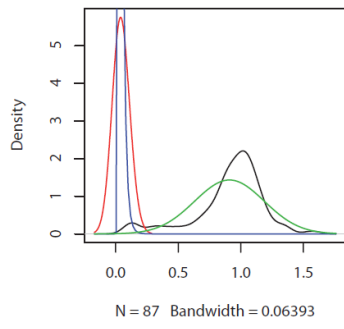
Q219



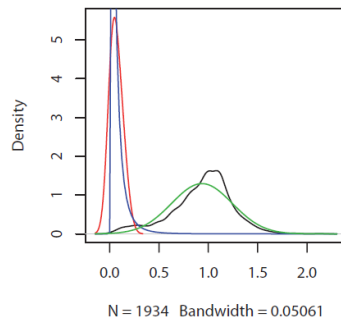
H225-2



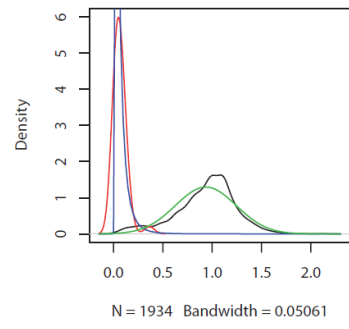
P225-2



L230-3



L236-2



Q8-PR-2

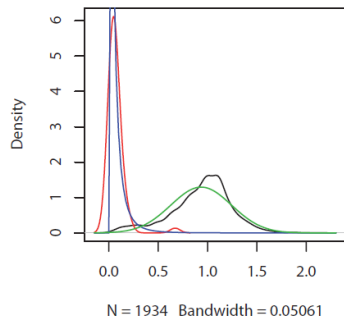
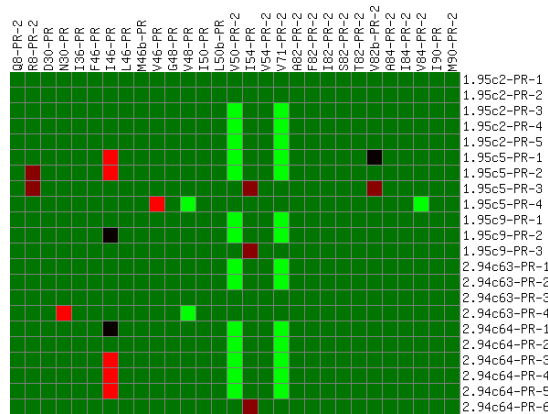


Figure H. Classification accuracy for the pure clonal samples belonging to the training set. Columns, probes included in the final version of the microarray that passed the quality control; Rows, samples hybridized to successive versions of the microarray. a) Hybridization of the PR region of the samples; b) Hybridization of the RT region; c) Hybridization of amplicons including both the PR and RT regions. Color code: Dark green, correctly classified signal (TP or TN); Dark red, FN signal; Red, FP signal; Black, UD signal; Light green, no data (due to spots discarded during the quality control or to the absence of certain probes in different versions of the microarray).

a)



b)



c)

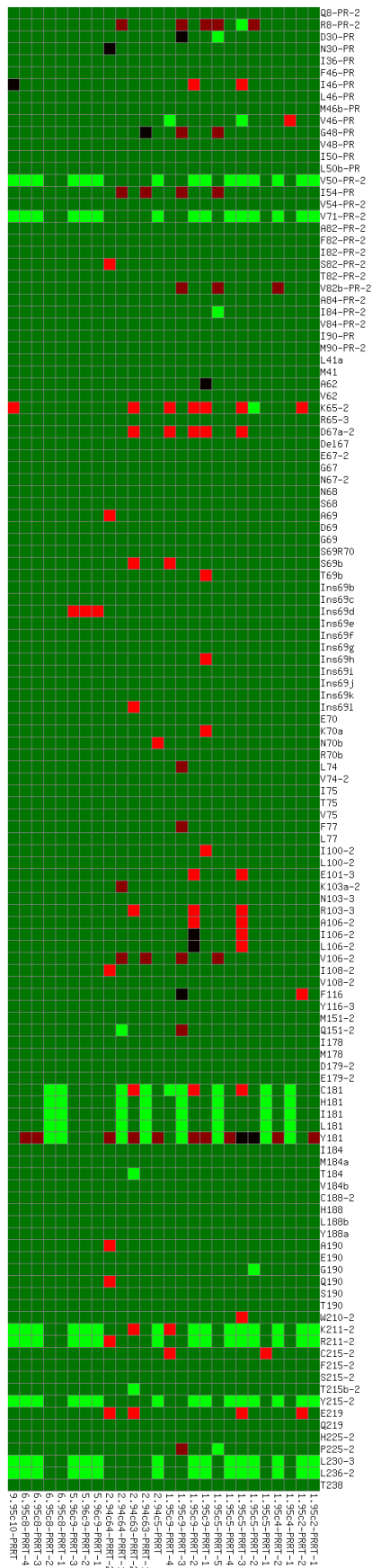
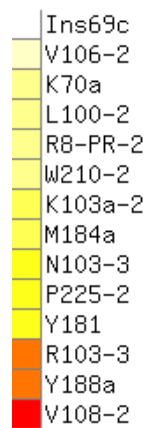


Figure I. Accumulation of errors (FP+FN+UD) during classification of clinical samples.

a) Errors per probe (bar: 11 to 20 errors); b) Errors per sample (bar: 10 to 22 errors). The remaining probes and hybridized samples showed a lower number of errors.

a)



b)

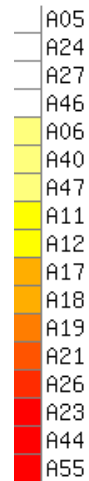
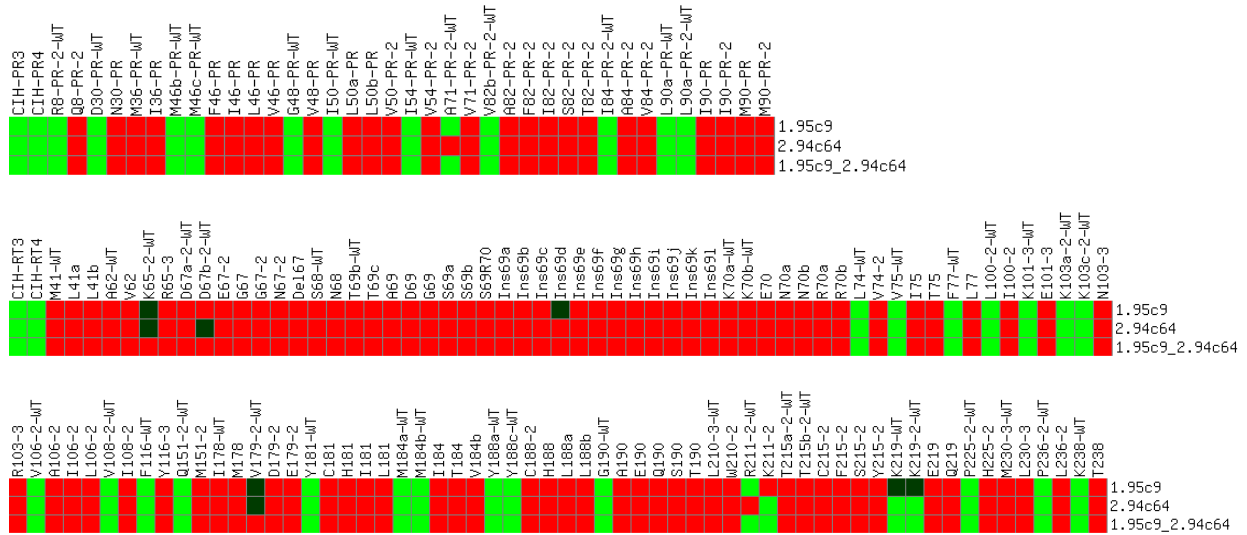


Figure J. Theoretical hybridization tables of binary mixtures of clonal samples belonging to the training set. a) Samples 1.95c9, 2.94c64 and their mixture; b) Samples pWT, pINS and their mixture. Color code: Green, expected hybridization; Red, not expected hybridization; Black, partial hybridization (one mismatch between probe and target is present at the 5' or 3' nucleotide of the hybridizing sequence).

a)



b)

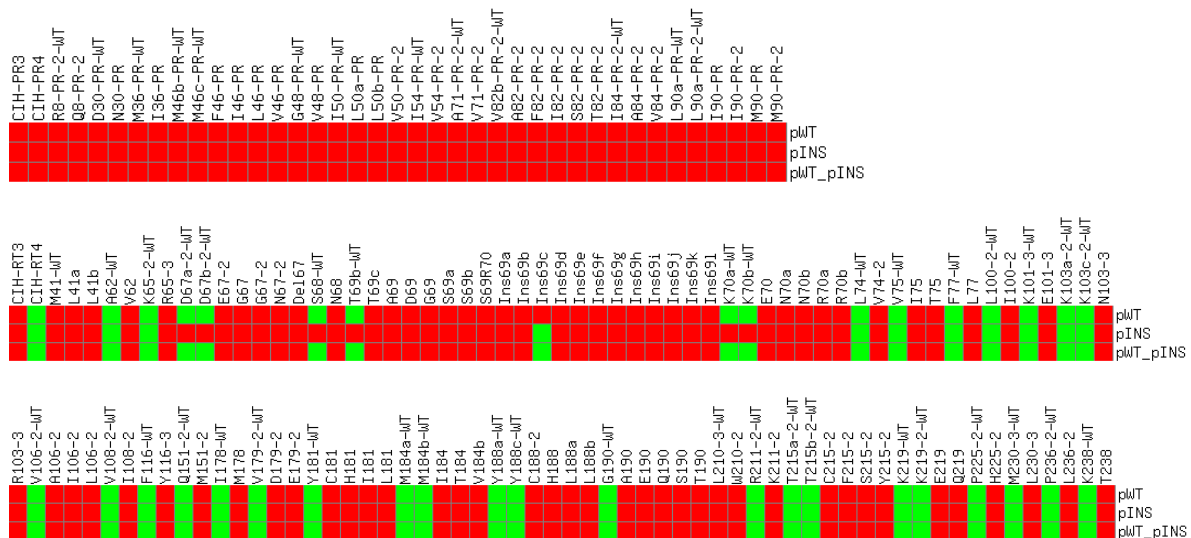
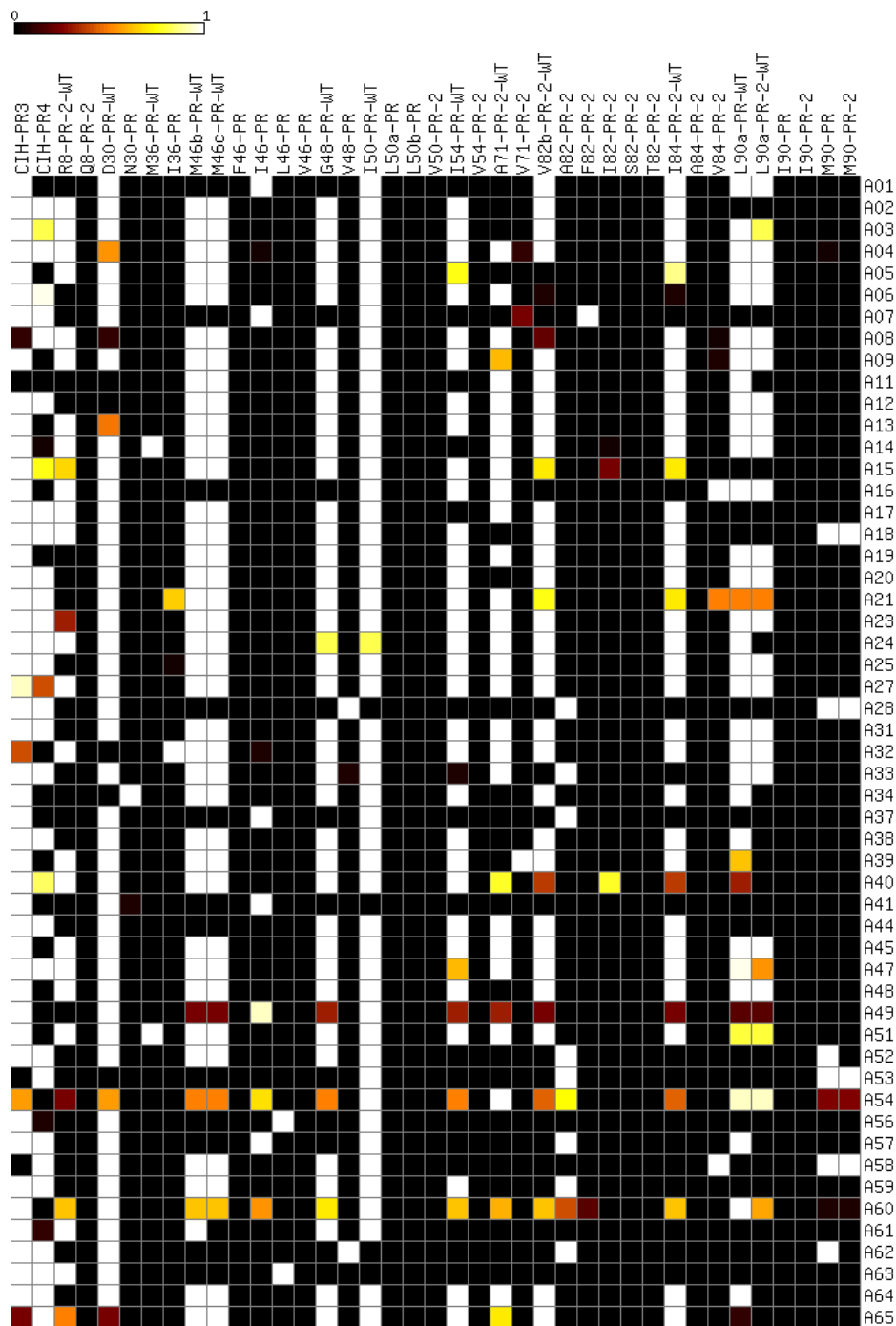


Figure K. Proportion of variants present in the clinical samples of the test set, determined by clonal sequencing. a) Probes of the PR region (columns) vs. sequences of the 53 target molecules that could be amplified out of the 57 clinical samples (rows). b) Probes of the RT region vs. sequences of the 51 amplified RT clinical samples. Bar: fraction of codons containing each queried codon within the quasiespecies. c) Distribution of variant proportions among the target samples.

a)



b)



c)

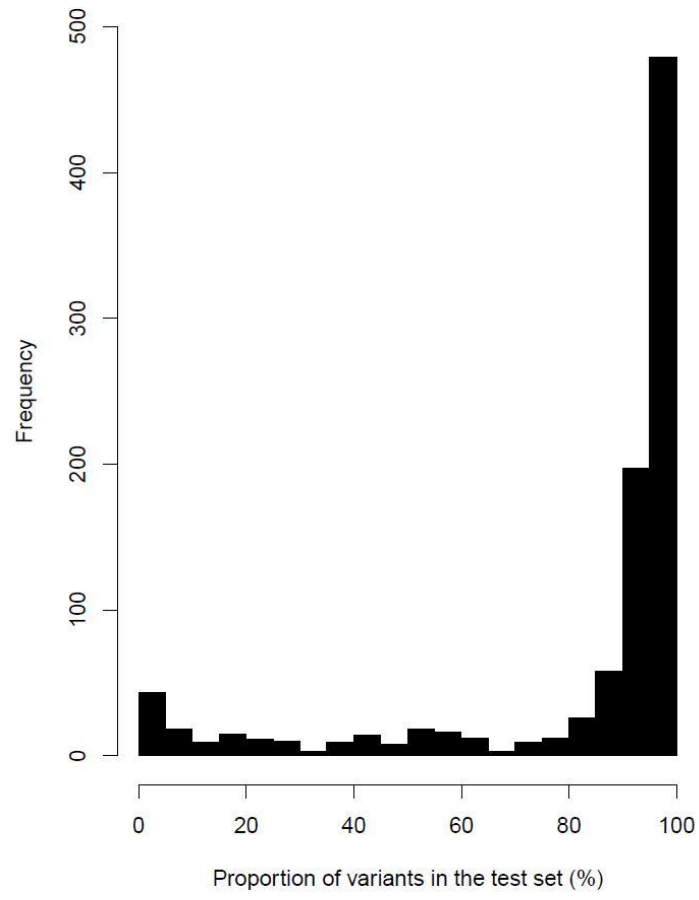


Figure L. Theoretical hybridization tables of the test set, using a preliminary detection threshold of 10% for minority sequences. Only perfect probe-target hybridizations are allowed. Columns, queried codons included in probes belonging to the PR (a) or RT (b) region; Rows, sequence of the quasiespecies of each clinical sample (see legend of Supp. Fig. S11). Color code: Green, expected hybridization (>10% of the clonal sequences of the sample match with the probe); Black, residual hybridization (probe matching in 5-10% of the sample clones); Red, not expected hybridization.

a)



b)



Table A. Sequences of the oligonucleotide probes printed onto the microarray for the screening of HIV-1 PR and RT drug and multidrug-resistance mutations. Discriminating probes are shown in black, positive HIV-1 internal hybridization controls (IHC-PR3, IHC-PR4, IHC-RT3 and IHC-RT4, spanning codons 25-29 and 94-99 of the PR, as well as 23-28 and 167-171 of the RT, respectively) are depicted in green, and negative controls (FMDV-G142-15r and FMDV-E142-15r, corresponding to sequences of the foot-and-mouth disease virus VP1 coding region) are shown in red. Yellow background highlights the probes complementary to wild type sequences at each queried codon. Two probes containing degenerated nucleotides (Ins69k and Ins69l) that could hybridize with a number of different targets showing 6 nt insertions between RT codons 69 and 70 have been included. The 124 highly performing probes that were maintained after the quality control protocol (see main text and Supp. Fig. S7) are underlined. In the second column, the nucleotide mutation(s) with respect to the corresponding wild type sequence are marked in bold. The calculated, salt-adjusted melting temperatures (quantified by means of the software available at <http://www.basic.northwestern.edu/biotools/oligocalc.html>) are shown.

Oligo name	Nucleotide sequence (5'-3')	T _m (°C)
IHC-PR3	TTTTTTTTTTTTTTTTGATACAGGAGCAGAT	54
IHC-PR4	TTTTTTTTTTTTTTTTGTTGCACTTTAAATTTT	51
R8-PR-2	TTTTTTTTTTTTTTTTTGCAACGACCCCTC	55
Q8-PR-2	TTTTTTTTTTTTTTTTTGCAAC A ACCCCTC	53
D30-PR	TTTTTTTTTTTTTTTTTAGCAGATGATACAGT	51
N30-PR	TTTTTTTTTTTTTTTTTAGCAGAT A AATACAGT	49
M36-PR	TTTTTTTTTTTTTTTTTAAGAAATGAGTTTGC	50
I36-PR	TTTTTTTTTTTTTTTTTAAGAAAT A AGTTTGC	49
M46b-PR	TTTTTTTTTTTTTTTTTCAAAAATGATAGGGG	51
M46c-PR	TTTTTTTTTTTTTTTTTACCAAAAATGATAGG	49
F46-PR	TTTTTTTTTTTTTTTTTCCAAAAT T CATAGGG	51
I46-PR	TTTTTTTTTTTTTTTTTCAAAAAT A ATAGGGG	49
L46-PR	TTTTTTTTTTTTTTTTTACCAAAAT T GATAGG	49
V46-PR	TTTTTTTTTTTTTTTTTACCAAAAT G TGATAGG	51
G48-PR	TTTTTTTTTTTTTTTTTATGATAGGGGAATT	51
V48-PR	TTTTTTTTTTTTTTTTTATGATAG T GGGAATT	49
I50-PR	TTTTTTTTTTTTTTTTTGGGGAATTGGAG	49
L50a-PR	TTTTTTTTTTTTTTTTTGGGGA T TAGGAG	49
L50b-PR	TTTTTTTTTTTTTTTTTGGGGA C TGGAG	49
V50-PR-2	TTTTTTTTTTTTTTTTTGGTGA G TGGAG	49
I54-PR	TTTTTTTTTTTTTTTTTAGGTTTTATCAAAGTA	49
V54-PR-2	TTTTTTTTTTTTTTTTTAGGATT G TCAAAGTA	51
A71-PR-2	TCCTCCTCCTCCTCCATAAAGCTATAGGTA	51
V71-PR-2	TCCTCCTCCTCCTCCATAAAG T TATAGGTA	49
V82b-PR-2	TCCTCCTCCTCCTCCTACACCTGTCAACAT	51
A82-PR-2	TCCTCCTCCTCCTCCACACCT G CCAACAT	52
F82-PR-2	TCCTCCTCCTCCTCCTACAC C TTCAACAT	49

<u>I82-PR-2</u>	TCCTCCTCCTCCTCCTACACCT A TCAACAT	49
<u>S82-PR-2</u>	TCCTCCTCCTCCTCCTACACCT T CCAACAT	51
<u>T82-PR-2</u>	TCCTCCTCCTCCTCCTACACCT A CCAACAT	51
I84-PR-2	TCCTCCTCCTCCTCCTGTCAACATAATTGG	49
<u>A84-PR-2</u>	TCCTCCTCCTCCTCCGTCAAC G CAATTGG	52
<u>V84-PR-2</u>	TCCTCCTCCTCCTCCGTCAAC G TAATTGG	49
L90a-PR	TTTTTTTTTTTTTTTTAATCTGTTGACTCAG	51
L90a-PR-2	TTTTTTTTTTTTTTTTAATCTGTTGACTCAGA	54
<u>I90-PR</u>	TTTTTTTTTTTTTTTTAATCTG A TAACTCAG	49
<u>I90-PR-2</u>	TTTTTTTTTTTTTTTTAATCTG A TAACTCAGA	51
<u>M90-PR</u>	TTTTTTTTTTTTTTTTAAATCTG A TGACTCA	49
<u>M90-PR-2</u>	TTTTTTTTTTTTTTTTAATCTG A TGACTCAGA	54
<hr/>		
IHC-RT3	TTTTTTTTTTTTTTTTATGGCCATTGACAGA	54
IHC-RT4	TTTTTTTTTTTTTTTTATCTTAGAGCCTTTTA	51
M41	TTTTTTTTTTTTTTTTTACAGAGATGGAAA	49
<u>L41a</u>	TTTTTTTTTTTTTTTTTACAGAG T GGAAA	49
<u>L41b</u>	TTTTTTTTTTTTTTTTTACAGAG C TGGAAA	51
A62	TTTTTTTTTTTTTTTTTAGTATTTGCCATAAAG	51
<u>V62</u>	TTTTTTTTTTTTTTTTTAGTATTTG T CATAAAG	49
K65-2	TCCTCCTCCTCCTCCCATAAAGAAAAAAGAC	49
<u>R65-3</u>	TCCTCCTCCTCCTCCCATAAAG A ATAGAC	51
D67a-2	TCCTCCTCCTCCTCCAAAAAAGACAGTACTA	49
D67b-2	TCCTCCTCCTCCTCCGAAAAAAGACAGTACT	51
<u>E67-2</u>	TCCTCCTCCTCCTCCAAAAAAG A AGTACTAA	49
<u>G67</u>	TTTTTTTTTTTTTTTTTAAAAAAG G CAGTACTA	51
<u>G67-2</u>	TCCTCCTCCTCCTCCAAAAAAG G CAGTACTA	51
<u>N67-2</u>	TCCTCCTCCTCCTCCGAAAAA A ACAGTACT	49
<u>Del67</u>	TTTTTTTTTTTTTTTTTAAAGAAAA A AGTACTAA	49
S68	TTTTTTTTTTTTTTTTTAAAAGACAGTACTAAAT	51
<u>N68</u>	TTTTTTTTTTTTTTTTTAAAAGACA A TACTAAAT	49
T69b	TTTTTTTTTTTTTTTTTGACAGTACTAAATGG	51
T69c	TTTTTTTTTTTTTTTTTGACAGTAC A AAATGG	51
<u>A69</u>	TTTTTTTTTTTTTTTTTAGACAGT G CTAAATG	51
<u>D69</u>	TTTTTTTTTTTTTTTTTGACAGT G AATAATGG	51
<u>G69</u>	TTTTTTTTTTTTTTTTTGACAGT G GTAATG	49
<u>S69a</u>	TTTTTTTTTTTTTTTTTGACAGT A GTAATGG	51
<u>S69b</u>	TTTTTTTTTTTTTTTTTGACAGT A GAAAATGG	51
<u>S69R70</u>	TTTTTTTTTTTTTTTTTACAGT T CTAGATGGA	51
<u>Ins69a</u>	TTTTTTTTTTTTTTTTT A G T A G T A G T AGTAAAT	49
<u>Ins69b</u>	TTTTTTTTTTTTTTTTT G T T C T A G T T C T AGAT	49
<u>Ins69c</u>	TTTTTTTTTTTTTTTTT C G A G T A G T T C G AAA	51
<u>Ins69d</u>	TTTTTTTTTTTTTTTTT A G T A G T A G T G G TAAA	49
<u>Ins69e</u>	TTTTTTTTTTTTTTTTT A G T G G T A G A T G G	49
<u>Ins69f</u>	TTTTTTTTTTTTTTTTT A G T A G T A G C G C TAAA	51
<u>Ins69g</u>	TTTTTTTTTTTTTTTTT A C T A G C A G C G C T A	51
<u>Ins69h</u>	TTTTTTTTTTTTTTTTT A G T A G T A G A G C AAAA	49
<u>Ins69i</u>	TTTTTTTTTTTTTTTTT A G T T C T A C T C T A G A	51

<u>Ins69j</u>	TTTTTTTTTTTTTTTTTAGTAGCGTGACTAAA	51
<u>Ins69k</u>	TTTTTTTTTTTTTTTTTAGTACTAGTDSTAAAT	49-51
<u>Ins69l</u>	TTTTTTTTTTTTTTTTTGTWSTAGTDSTARAT	46-51
K70a	TTTTTTTTTTTTTTTTTCAGTACTAAATGGAG	51
K70b	TTTTTTTTTTTTTTTTTGTACTAAATGGAGAA	49
<u>E70</u>	TTTTTTTTTTTTTTTTTCAGTACTGAATGGA	49
<u>N70a</u>	TTTTTTTTTTTTTTTTTGTACTAAATGGAGAA	49
<u>N70b</u>	TTTTTTTTTTTTTTTTTGTACTAACTGGAGAA	51
<u>R70a</u>	TTTTTTTTTTTTTTTTTAGTACTAGATGGAGA	51
<u>R70b</u>	TTTTTTTTTTTTTTTTTGTACTAGGTGGAG	51
L74	TTTTTTTTTTTTTTTTTGAGAAAATTAGTAGAT	49
<u>V74-2</u>	TTTTTTTTTTTTTTTTTGAGAATAGTAGTAGAT	51
V75	TTTTTTTTTTTTTTTTTGAAAATTAGTAGATTTTC	51
<u>I75</u>	TTTTTTTTTTTTTTTTTGAAAATTAATAGATTTTC	49
<u>T75</u>	TTTTTTTTTTTTTTTTTGAAAATTAACAGATTTTC	51
F77	TTTTTTTTTTTTTTTTTAGTAGATTTTCAGAGA	49
<u>L77</u>	TTTTTTTTTTTTTTTTTAGTAGATCTCAGAGA	51
L100-2	TCCTCCTCCTCCTCCCGCAGGGTTAAAAAA	51
<u>I100-2</u>	TCCTCCTCCTCCTCCCGCAGGGATAAAAAA	51
K101-3	TCCTCCTCCTCCTCCGCAGGGTTAAAAAAGA	54
<u>E101-3</u>	TCCTCCTCCTCCTCCGCAGGGTTAGAAAAGA	54
K103a-2	TCCTCCTCCTCCTCCAAAAAGAAAAAATCAGT	49
K103c-2	TCCTCCTCCTCCTCCAAAAAGAAAAAATCAG	49
<u>N103-3</u>	TCCTCCTCCTCCTCCAAATAGAACAAATCAGT	51
<u>R103-3</u>	TCCTCCTCCTCCTCCAAATAGAGAAAATCAG	51
V106-2	TCCTCCTCCTCCTCCAAATCAGTAACAGTA	49
<u>A106-2</u>	TCCTCCTCCTCCTCCAAATCAGCAACAGTA	51
<u>I106-2</u>	TCCTCCTCCTCCTCCAAAAATCAATAACAGTA	49
<u>L106-2</u>	TCCTCCTCCTCCTCCAAAAATCATTAACAGTA	49
V108-2	TTTTTTTTTTTTTTTTTCAGTAACAGTACTGG	54
<u>I108-2</u>	TTTTTTTTTTTTTTTTTCAGTAACAATACTGG	51
F116	TTTTTTTTTTTTTTTTTGCATATTTTTCAGTTC	51
<u>Y116-3</u>	TTTTTTTTTTTTTTTTTGCATATTAATCACTTC	51
Q151-2	TTTTTTTTTTTTTTTTTGCTTCCACAGGGAT	55
<u>M151-2</u>	TTTTTTTTTTTTTTTTTGCTTCCAATGGGAT	52
I178	TTTTTTTTTTTTTTTTTCAGACATAGTTATCT	49
<u>M178</u>	TTTTTTTTTTTTTTTTTCAGACATGGTTATCT	51
V179-2	TTTTTTTTTTTTTTTTTAGACATAGTTATCTATC	53
<u>D179-2</u>	TTTTTTTTTTTTTTTTTAGACATAGATATCTATC	53
<u>E179-2</u>	TTTTTTTTTTTTTTTTTAGACATAGAGATCTAT	51
Y181	TTTTTTTTTTTTTTTTTAGTTATCTATCAATAC	49
<u>C181</u>	TTTTTTTTTTTTTTTTTAGTTATCTGTCAATAC	51
<u>H181</u>	TTTTTTTTTTTTTTTTTAGTTATCCATCAATAC	51
<u>I181</u>	TTTTTTTTTTTTTTTTTAGTTATCATCAATAC	49
<u>L181</u>	TTTTTTTTTTTTTTTTTAGTTATCCTCAATAC	51
M184a	TTTTTTTTTTTTTTTTTCAATACATGGATGATT	51
M184b	TTTTTTTTTTTTTTTTTCAATACATGGATGAT	49

<u>I184</u>	TTTTTTTTTTTTTTTCAATACAT A GATGATT	49
<u>T184</u>	TTTTTTTTTTTTTTTCAATACAC G GATGAT	51
<u>V184b</u>	TTTTTTTTTTTTTTTCAATAC G T A GATGAT	49
<u>Y188a</u>	TTTTTTTTTTTTTTTGATTTGTATGTAGGA	49
<u>Y188c</u>	TTTTTTTTTTTTTTTGATTTGTATGTAGGAT	51
<u>C188-2</u>	TTTTTTTTTTTTTTTGATTAGT G TGTAGGA	51
<u>H188</u>	TTTTTTTTTTTTTTTGATTT G CATGTAGG	51
<u>L188a</u>	TTTTTTTTTTTTTTTGATTTGT T AGTAGGAT	51
<u>L188b</u>	TTTTTTTTTTTTTTTGATTT G C T TGTAGG	51
<u>G190</u>	TTTTTTTTTTTTTTTATGTAGGATCTGAC	51
<u>A190</u>	TTTTTTTTTTTTTTTATGTAG C ATCTGAC	51
<u>E190</u>	TTTTTTTTTTTTTTTATGTAG A ATCTGAC	49
<u>Q190</u>	TTTTTTTTTTTTTTTATGTAC A ATCTGAC	49
<u>S190</u>	TTTTTTTTTTTTTTTATGT A AGCTCTGAC	51
<u>T190</u>	TTTTTTTTTTTTTTTATGTAA C ATCTGAC	49
<u>L210-3</u>	TTTTTTTTTTTTTTTATCTGTTGAGTTGGG	54
<u>W210-2</u>	TTTTTTTTTTTTTTTATCTGT G GAGTTGG	52
<u>R211-2</u>	TTTTTTTTTTTTTTTGTTGAGGTGGGGA	53
<u>K211-2</u>	TTTTTTTTTTTTTTTGTTGA A GTGGGGA	51
<u>T215a-2</u>	TTTTTTTTTTTTTTTGGACTTACCACACC	55
<u>T215b-2</u>	TTTTTTTTTTTTTTTGGGACTTACCACAC	55
<u>C215-2</u>	TTTTTTTTTTTTTTTGGACTT T GCACACC	55
<u>F215-2</u>	TTTTTTTTTTTTTTTGGACTT T T C ACACC	52
<u>S215-2</u>	TTTTTTTTTTTTTTTGGGACTT T CCACAC	55
<u>Y215-2</u>	TTTTTTTTTTTTTTTGGACTT T ACACACC	52
<u>K219</u>	TTTTTTTTTTTTTTTACCAGACAAAAACA	49
<u>K219-2</u>	TCCTCCTCCTCCTCCACCAGACAAAAACA	49
<u>E219</u>	TTTTTTTTTTTTTTTACCAGAC G AAAAACA	51
<u>Q219</u>	TTTTTTTTTTTTTTTACCAGAC C AAAAACA	51
<u>P225-2</u>	TCCTCCTCCTCCTCCAAGAACCTCCATTC	51
<u>H225-2</u>	TCCTCCTCCTCCTCCAAGAAC A TCCATTC	49
<u>M230-3</u>	TTTTTTTTTTTTTTTCTTTGGATGGGTTAT	51
<u>L230-3</u>	TTTTTTTTTTTTTTTCTTTG C TGGGTTAT	54
<u>P236-2</u>	TCCTCCTCCTCCTCCCTCCATCCTGATAAAT	54
<u>L236-2</u>	TCCTCCTCCTCCTCCCTCCAT C TGATAAAT	51
<u>K238</u>	TTTTTTTTTTTTTTTCTGATAAATGGAC	49
<u>T238</u>	TTTTTTTTTTTTTTTCTGATA C ATGGAC	51
FMDV-G142-15r	TTTTTTTTTTTTTTTCAAATCCCCGCGTG	56
FMDV-E142-15r	TTTTTTTTTTTTTTTCAAATCCTCGCGTG	54

Table B. Classification accuracy of binary mixtures of pure clonal samples used to set the preliminary detection limit of the microarray. Step 4 did not increase the accuracy in this series of experiments because none of the full hybridized microarrays had to be discarded.

Signals	Classification accuracy (% of signals) after filtering							
	Step 1: Spot filter		Step 2: Probe filter 1 (overlapped)		Step 3: Probe filter 2 (duplications)		Step 4: Array filter	
	1.95c9/ 2.94c64 (RT and PR-RT)	pWT/ pINS (RT)	1.95c9/ 2.94c64 (RT and PR-RT)	pWT/ pINS (RT)	1.95c9/ 2.94c64 (RT and PR-RT)	pWT/ pINS (RT)	1.95c9/ 2.94c64 (RT and PR-RT)	pWT/ pINS (RT)
Correct	95.32	91.15	97.80	93.08	97.64	94.26	97.64	94.26
FP	2.61	3.35	1.37	2.13	1.50	1.44	1.50	1.44
FN	1.44	4.77	0.49	3.93	0.54	3.36	0.54	3.36
UD	0.62	0.73	0.34	0.86	0.32	0.93	0.32	0.93

Table C. Minority subpopulations present in the clinical samples of the test set, quantified by means of clonal analysis.

Range of minority variants within the quasispecies (%)	Fraction of positions (PR + RT)
1.00 – 4.99	15/970 (1.55%)
5.00 – 9.99	34/970 (3.50%)
10.00 – 19.99	24/970 (2.47%)

