Supplementary Information

Title: Quantitative and sensitive detection of rare mutations using droplet-based microfluidics.

Authors: Deniz Pekin[1], Yousr Skhiri[1], Jean-Christophe Baret[1,3], Delphine Le Corre[2], Linas Mazutis[1], Chaouki Ben Salem[1], Florian Millot[1], Abdeslam El Harrak[4], J. Brian Hutchison[5], Jonathan W. Larson[5], Darren R. Link[5], Pierre Laurent-Puig[2], Andrew D. Griffiths[1,*] and Valérie Taly[1,*].

[1]Institut de Science et d'Ingénierie Supramoléculaires (ISIS); Université de Strasbourg; CNRS UMR 7006, 8 allée Gaspard Monge, BP 70028, F-67083 Strasbourg Cedex, France

[2]Université Paris Descartes; INSERM UMR-S775; Centre Universitaire des Saints-Pères, 45 rue des Saints-Pères, 75270 Paris Cedex 06, France

[3]Max-Planck-Institute for Dynamics and Self-organization; Am Fassberg 17, D-37077 Göttingen, Germany

[4]RainDance Technologies France; 8 allée Gaspard Monge, BP 70028, F-67083 Strasbourg Cedex, France

[5]RainDance Technologies; Lexington, MA 02421, Massachusetts, USA

[*]To whom correspondence should be addressed. (V.T.) Phone: +33 (0)368 855 213. Fax: +33 (0)368 855 115. e-mail: vtaly@unistra.fr. (A.D.G.) Phone: +33 (0) 368 855 171. Fax: +33 (0)368 855 115. e-mail: griffiths@unistra.fr.

Supplementary Text

Statistical analysis

Mixtures of wild-type (wt) and mutant (m) *KRAS* genes were compartmentalized in droplets. The ratio of mutant to wild-type *KRAS* genes was varied experimentally over 5 decades while the ratio of wild-type *KRAS* genes droplets over the total number of droplets was kept constant (0.08). Four types of droplets were then generated: red-fluorescent droplets containing wild-type *KRAS* genes, green-fluorescent droplets containing mutant *KRAS* genes, non-fluorescent droplets containing no *KRAS* genes and yellow droplets containing both wild-type and mutant genes as a result of the combination of green and red fluorescence. It has been shown that the encapsulation of single gene in droplets is described by a Poisson distribution.[1] The probability to encapsulate k genes in one droplet depends on the average number of genes per droplet λ as:

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!} \tag{S1}$$

The encapsulation of k_m mutant genes and k_{wt} wild-type genes in a single droplet is given by the product of the Poisson distribution of the two independent random variable k_m and k_{wt} (see also[2] for an analogous example).

$$p_{k_m,k_{wt}} = \frac{\lambda_m^{k_m} e^{-\lambda_m}}{k_m!} \times \frac{\lambda_{wt}^{k_{wt}} e^{-\lambda_{wt}}}{k_{wt}!} \tag{S2}$$

The statistics of encapsulation are presented in Supplementary Table S1 for clarity for three different cases (λ_{wt}= 0.8 (a); 0.08 (b) and 0.008 (c)). Analytically, the occurrence of the droplet color (green, red, yellow and black) can be derived from the Poisson distribution:

$$\frac{N_{\mathrm{g}}}{N} = e^{-\lambda_{wt}}(1-e^{-\lambda_m})$$

$$\frac{N_{\mathrm{r}}}{N} = e^{-\lambda_m}(1-e^{-\lambda_{wt}})$$

$$\frac{N_{\mathrm{y}}}{N} = (1-e^{-\lambda_{wt}})(1-e^{-\lambda_m})$$

$$\frac{N_b}{N} = e^{-(\lambda_m+\lambda_{wt})}$$

Where N_g, N_r, N_y and N_b are respectively the number of green, red, yellow and black (empty) droplets. N is the total number of droplets. It is clear that for high λ_{wt} (see Supplementary Table S1a) the number of yellow droplets have to be evaluated in order to count all the mutant and wild-type genes (see eq. 4 in the main text). When a very low λ_{wt} is used (λ_{wt}=0.008), the yellow droplets appear insignificant in the statistics (Supplementary Table S1c). However in this case, the experiment requires the analysis of a majority of empty droplets, which reduces the throughput of the test. Finally for intermediate values (λ_{wt} = 0.08, Table 1b), the yellow droplets represent less than 10% of the mutant droplet and have a minor impact on the analysis. In the following, we will focus on the three droplet types: empty, red and green.

Confidence intervals for the dilutions

When a sub-population of N droplets was analyzed we obtained N_r red droplets (wild-type) and N_g green droplets (mutant). The rest ($N - N_r - N_g$) were droplets that did not contain target DNA. From these measurements the ratio $F = N_g / N_r$ was determined and compared to the theoretical value, namely from the dilution of mutant into wild-type genes, $F^* = N_m / N_{wt}$. Statistically, the value F differs from F^*. In the following we discuss the confidence intervals for the determination of the ratio F^* based on the fraction F obtained by droplet counting.

Since we randomly pick N droplets, the problem is equivalent to basic opinion polls. The determination of the fraction of red-fluorescent (wild-type) droplets (p_{wt}) within the 95% confidence interval is:

$$p_{wt} = \frac{N_r}{N} \pm 1.96\sqrt{\frac{N_r}{N}\frac{1-N_r/N}{N}} \tag{S3}$$

Alternatively, the fraction of green-fluorescent (mutant) droplets (p_m) within the 95% confidence interval is:

$$p_m = \frac{N_g}{N} \pm 1.96\sqrt{\frac{N_g}{N}\frac{1-N_g/N}{N}} \tag{S4}$$

The 95% confidence interval for the ratio $F = p_m/p_{wt}$ is in the general case complicated to determine.[3] This can be understood qualitatively: if the value 0 is in the confidence interval of the p_{wt}, the confidence interval will diverge to infinity. However, in our case, 0 is not in the confidence interval and with the assumption that the standard error for p_{wt} is smaller than p_{wt} (which is the case for large values of N and $p_{wt} \sim 0.1$) then the confidence interval for the ratio can be expressed from the standard errors of p_{wt} and p_m.[3] This leads to Eq. S5 (See Supplementary Figure S2):

$$F = \frac{p_m}{p_{wt}} = F^*\left(1 \pm 1.96\sqrt{\frac{1-N_r/N}{N_r} + \frac{1-N_g/N}{N_g}}\right) \tag{S5}$$

In the case where the number of green (mutant) droplets is much smaller than both the number of red (wild-type) droplets and the total number of droplets, the 95% confidence interval is mainly determined by the smallest population and is in good approximation simply:

$$F = F^*(1 \pm \frac{1.96}{\sqrt{N_g}}) \tag{S6}$$

Supplementary References

1. L. Mazutis, A. F. Araghi, O. J. Miller, J. C. Baret, L. Frenz, A. Janoshazi, V. Taly, B. J. Miller, J. B. Hutchison, D. Link, A. D. Griffiths and M. Ryckelynck, *Anal Chem*, 2009, **81**, 4813-4821.
2. J. C. Baret, O. J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M. L. Samuels, J. B. Hutchison, J. J. Agresti, D. R. Link, D. A. Weitz and A. D. Griffiths, *Lab Chip*, 2009, **9**, 1850-1858.
3. H. Motulski, Intuitive biostatistics, *Oxford University Press*, 1995.
4. L. Mazutis, J. C. Baret and A. D. Griffiths, *Lab Chip*, 2009, **9**, 2665-2672.

SUPPLEMENTARY TABLES

(a) λ_{WT} = 0.8000, λ_M = 0.0800

		Wild-type 0	1	2	3	4
Mutant	0	4.1E-01	3.3E-01	1.3E-01	3.5E-02	7.1E-03
	1	3.3E-02	2.7E-02	1.1E-02	2.8E-03	5.7E-04
	2	1.3E-03	1.1E-03	4.2E-04	1.1E-04	2.3E-05
	3	3.5E-05	2.8E-05	1.1E-05	3.0E-06	6.0E-07
	4	7.1E-07	5.7E-07	2.3E-07	6.0E-08	1.2E-08

Total Empty:	4.1E-01
Total Red:	5.1E-01
Total Green:	3.5E-02
Total Yellow:	4.2E-02
Total :	0.99858866
yellow/green:	1.22240000

(b) λ_{WT} = 0.0800, λ_M = 0.0080

		Wild-type 0	1	2	3	4
Mutant	0	9.2E-01	7.3E-02	2.9E-03	7.8E-05	1.6E-06
	1	7.3E-03	5.9E-04	2.3E-05	6.3E-07	1.3E-08
	2	2.9E-05	2.3E-06	9.4E-08	2.5E-09	5.0E-11
	3	7.8E-08	6.3E-09	2.5E-10	6.7E-12	1.3E-13
	4	1.6E-10	1.3E-11	5.0E-13	1.3E-14	2.7E-16

Total Empty:	9.2E-01
Total Red:	7.6E-02
Total Green:	7.4E-03
Total Yellow:	6.1E-04
Total :	0.99999997
yellow/green:	0.08328704

(c) λ_{WT} = 0.0080, λ_M = 0.0008

		Wild-type 0	1	2	3	4
Mutant	0	9.9E-01	7.9E-03	3.2E-05	8.5E-08	1.7E-10
	1	7.9E-04	6.3E-06	2.5E-08	6.8E-11	1.4E-13
	2	3.2E-07	2.5E-09	1.0E-11	2.7E-14	5.4E-17
	3	8.5E-11	6.8E-13	2.7E-15	7.2E-18	1.4E-20
	4	1.7E-14	1.4E-16	5.4E-19	1.4E-21	2.9E-24

Total Empty:	9.9E-01
Total Red:	8.0E-03
Total Green:	7.9E-04
Total Yellow:	6.4E-06
Total :	1.00000000
yellow/green:	0.00803209

Supplementary Table S1 Probabilities of occurrence of the different cases in the encapsulation of mutant and wild-type genes for different values of λ_{wt} and a constant ratio λ_m / λ_{wt}. The background color shows the resulting fluorescence of the droplets. Droplets containing no gene appear black, wild-type gene red and mutant gene green. When one or more mutant genes are coencapsulated with one or more wild-type genes the droplet appears yellow. See Supplementary Eq. S2 for the details of the calculations. From a) to c) the occupancy is decreasing resulting in a decrease of the occurrence of yellow droplets.

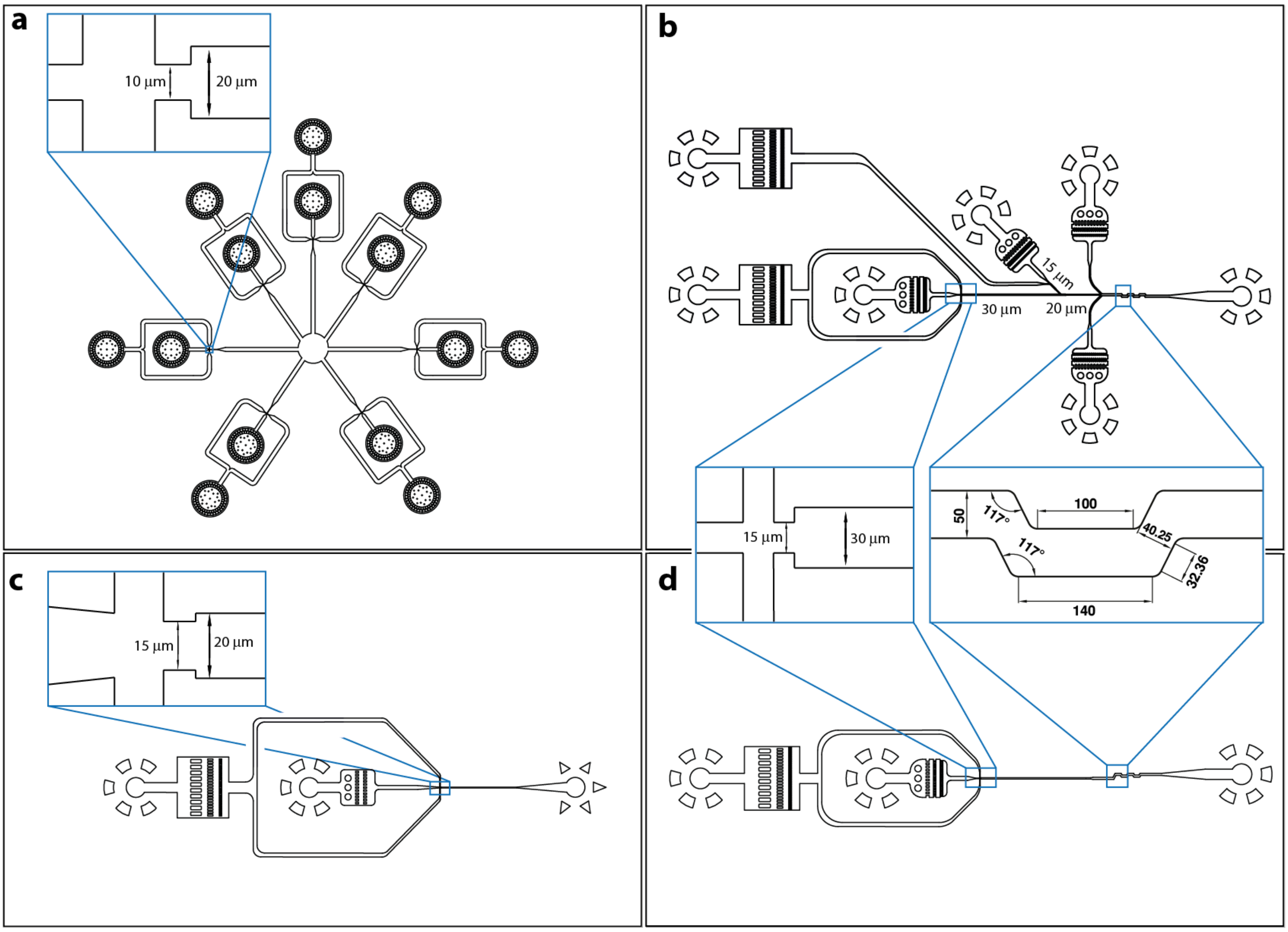
SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1: Microfluidic devices. (a) The device for producing seven different types of droplets, used for producing the probe emulsion for the parallel multiple mutation analysis. (b) The device used for passive droplet fusion.[4] (c) The device used for reinjection and on-chip droplet fluorescence analysis. (d) The device used for producing droplets on-chip, for sensitivity and MASI experiments.

Supplementary Figure S2. 95% confidence intervals determined from the measurement of *N* droplets. $N = 10^3$, 10^5 and 10^7 ($N_{wt} / N = 0.1$ for all cases). For a given ratio of mutant to wild-type genes, F^*, the ratio of the number of green droplets over the number of red droplet is equal to F^* if an infinite number of droplets is analyzed (dashed lines). The 95% confidence interval (orange region) is a function of the number of droplets N analyzed (Eq. S6). As expected, the accuracy of the measurement of the dilution increases with the number of droplets analyzed.

SUPPLEMENTARY FIGURES

Supplementary Figure S1.



Supplementary Figure S2.