Supplementary Information

A versatile genome-scale PCR-based pipeline for high-definition DNA FISH

Magda Bienko[1,*], Nicola Crosetto[1,*], Leonid Teytelman[1], Sandy Klemm[2], Shalev Itzkovitz[1], Alexander van Oudenaarden[1,3,§]

[1]Departments of Physics and Biology, and the Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

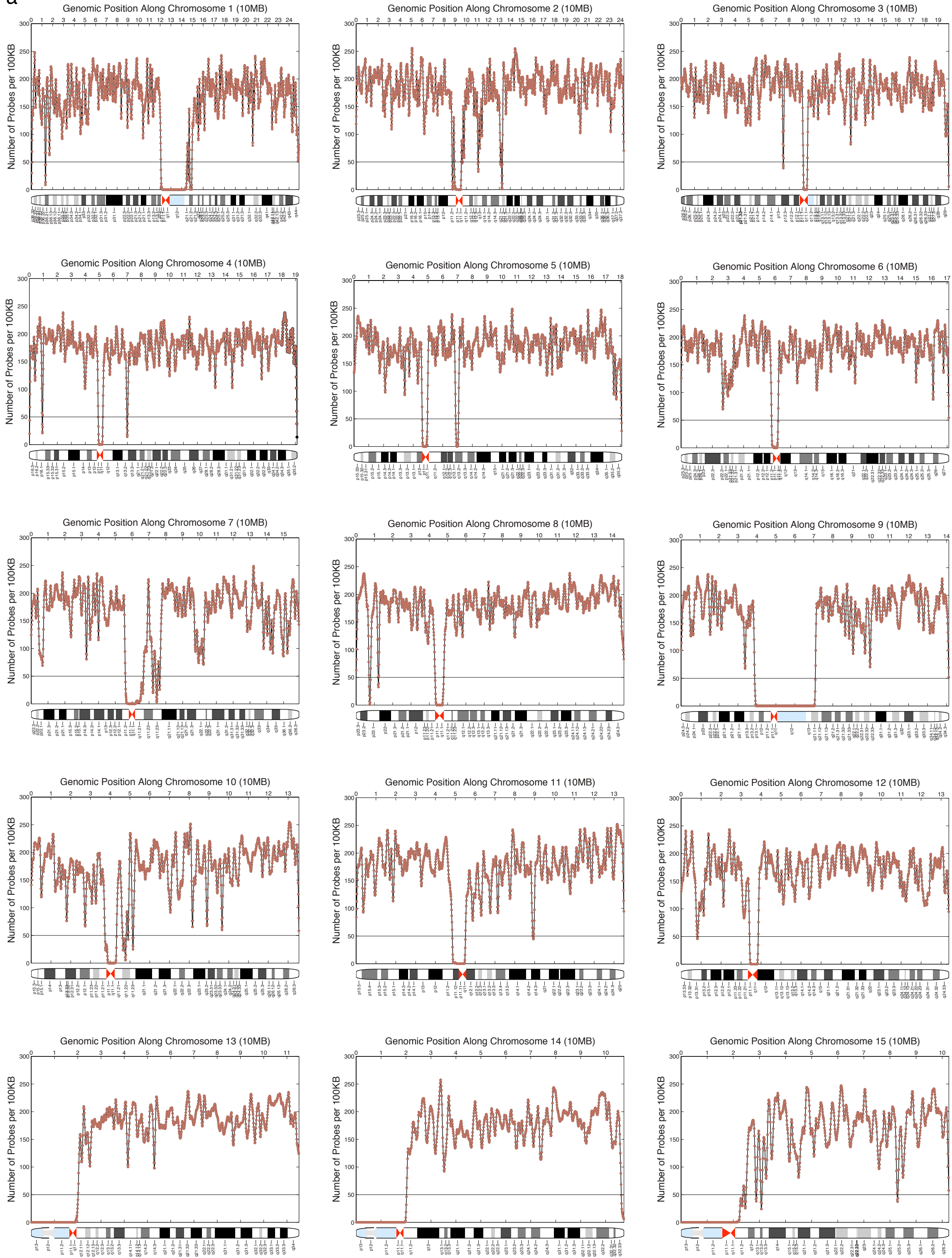
[2]Departments of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[3]Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, Netherlands.

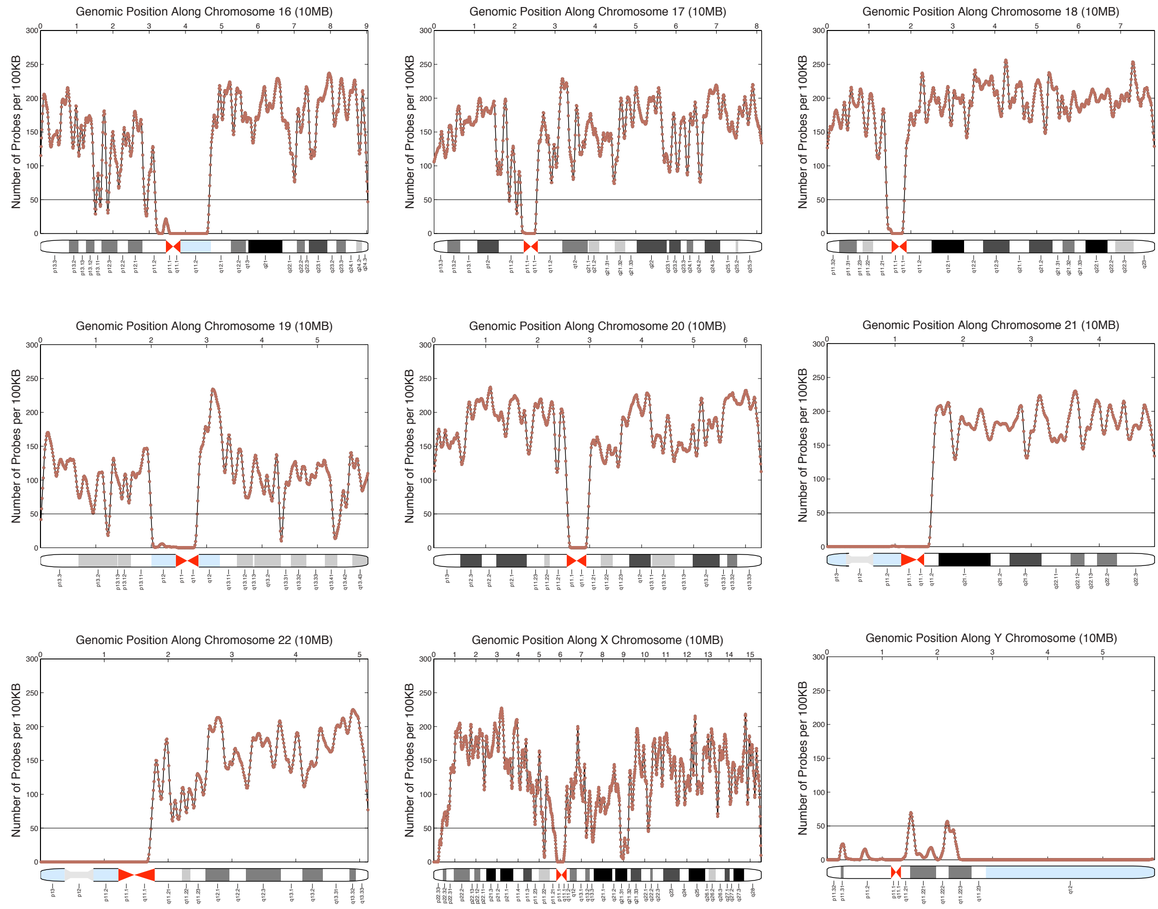
[§] Correspondence should be sent to A. v. O. (a.vanoudenaarden@hubrecht.eu)

[*] Equally contributing authors

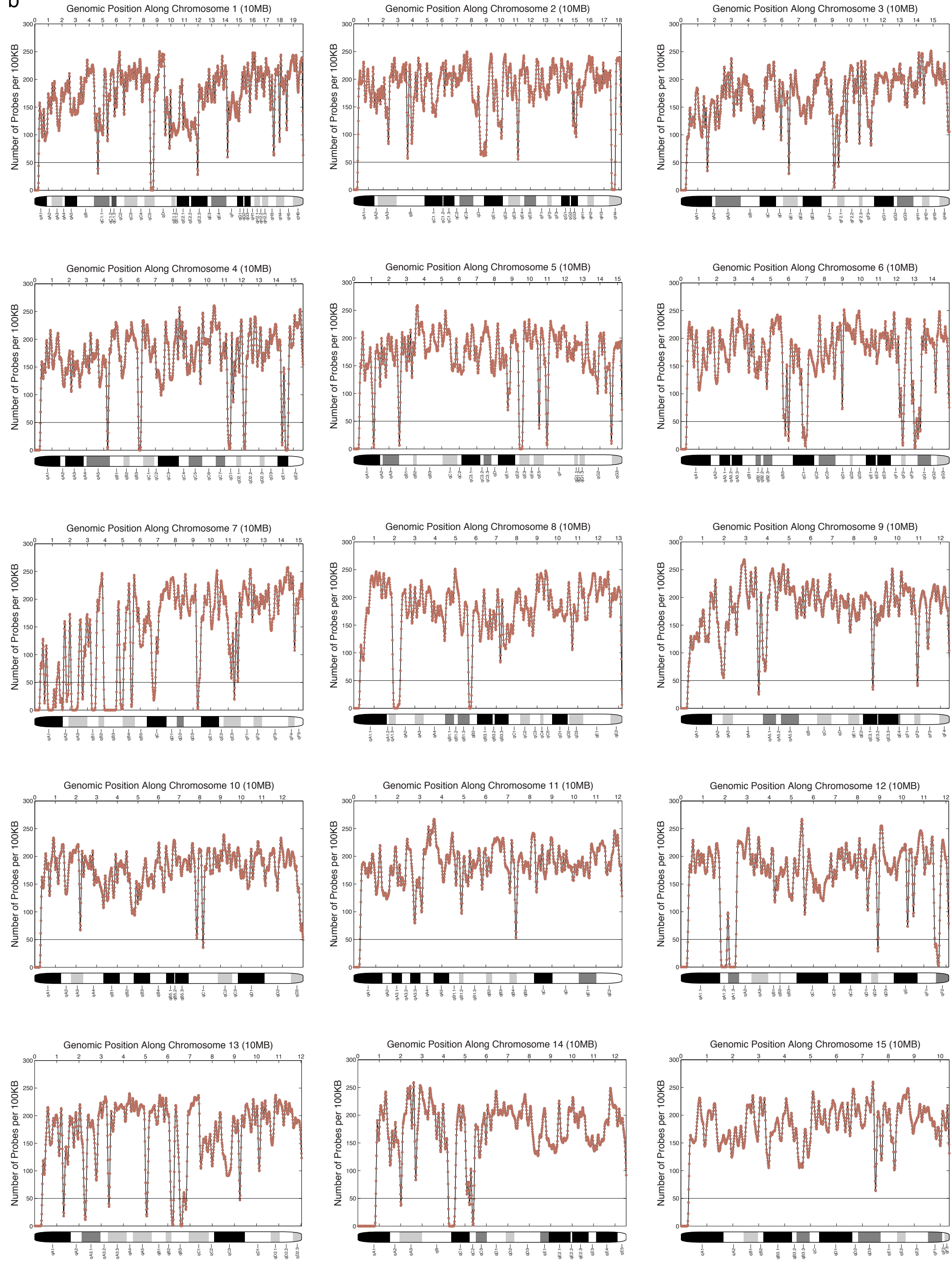
Supplementary Item	Title or Caption
Supplementary Figure 1	Distribution of HD-FISH amplicons along the human and mouse genomes
Supplementary Figure 2	HD-FISH cost analysis
Supplementary Figure 3	Distribution of spot counts for HER2 probes of decreasing size
Supplementary Figure 4	Spotting of chromosome 1 and 17 using HD-FISH probes
Supplementary Figure 5	Frequency distribution of HER2 mRNA counts
Supplementary Note	
Supplementary Video 1	3D rendering of Chr17 in HME cells, visualized with ten HD-FISH probes evenly spaced every 8 Mb and labeled with two alternating fluorophores (green: AlexaFluor594; magenta: AlexaFluor647). The nucleus displayed is the same as in the Z-projection shown in **Figure 3a** (mid panel).
Supplementary Video 2	3D animation of Chr17 in HME cells, visualized with sixteen HD-FISH probes spaced evenly every 5 Mb and labeled with two alternating fluorophores (green: AlexaFluor594; magenta: AlexaFluor647) together with a Chr17 paint probe (blue).


a
Genomic Position Along Chromosome 1 (10MB)
Genomic Position Along Chromosome 2 (10MB)
Genomic Position Along Chromosome 3 (10MB)
Genomic Position Along Chromosome 4 (10MB)
Genomic Position Along Chromosome 5 (10MB)
Genomic Position Along Chromosome 6 (10MB)
Genomic Position Along Chromosome 7 (10MB)
Genomic Position Along Chromosome 8 (10MB)
Genomic Position Along Chromosome 9 (10MB)
Genomic Position Along Chromosome 10 (10MB)
Genomic Position Along Chromosome 11 (10MB)
Genomic Position Along Chromosome 12 (10MB)
Genomic Position Along Chromosome 13 (10MB)
Genomic Position Along Chromosome 14 (10MB)
Genomic Position Along Chromosome 15 (10MB)
Number of Probes per 100KB

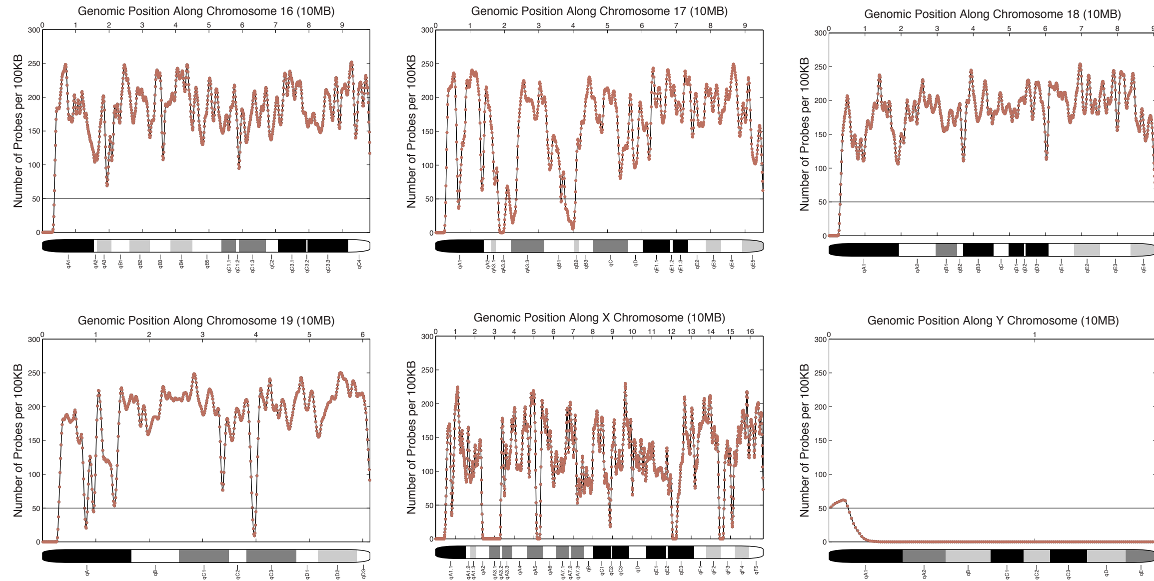
a



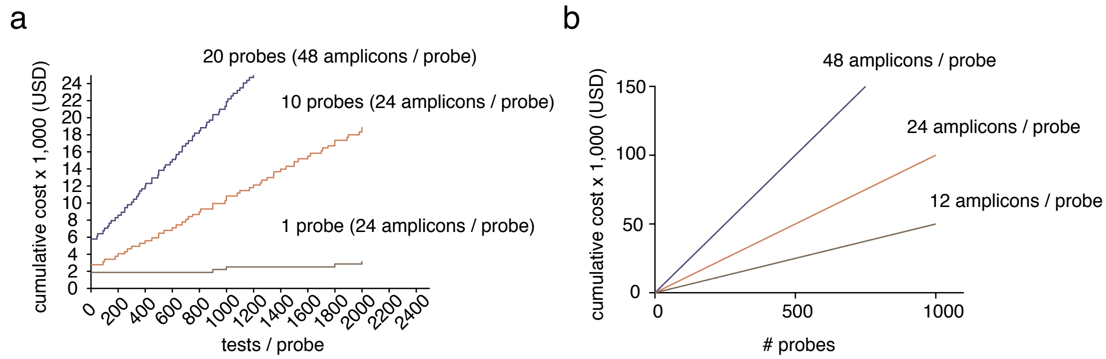
b



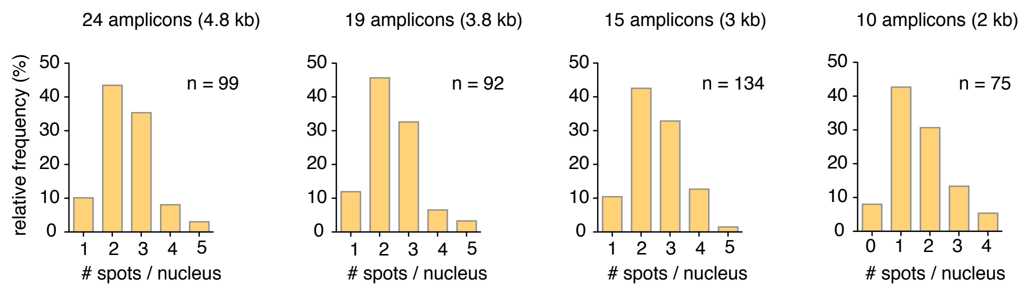
b



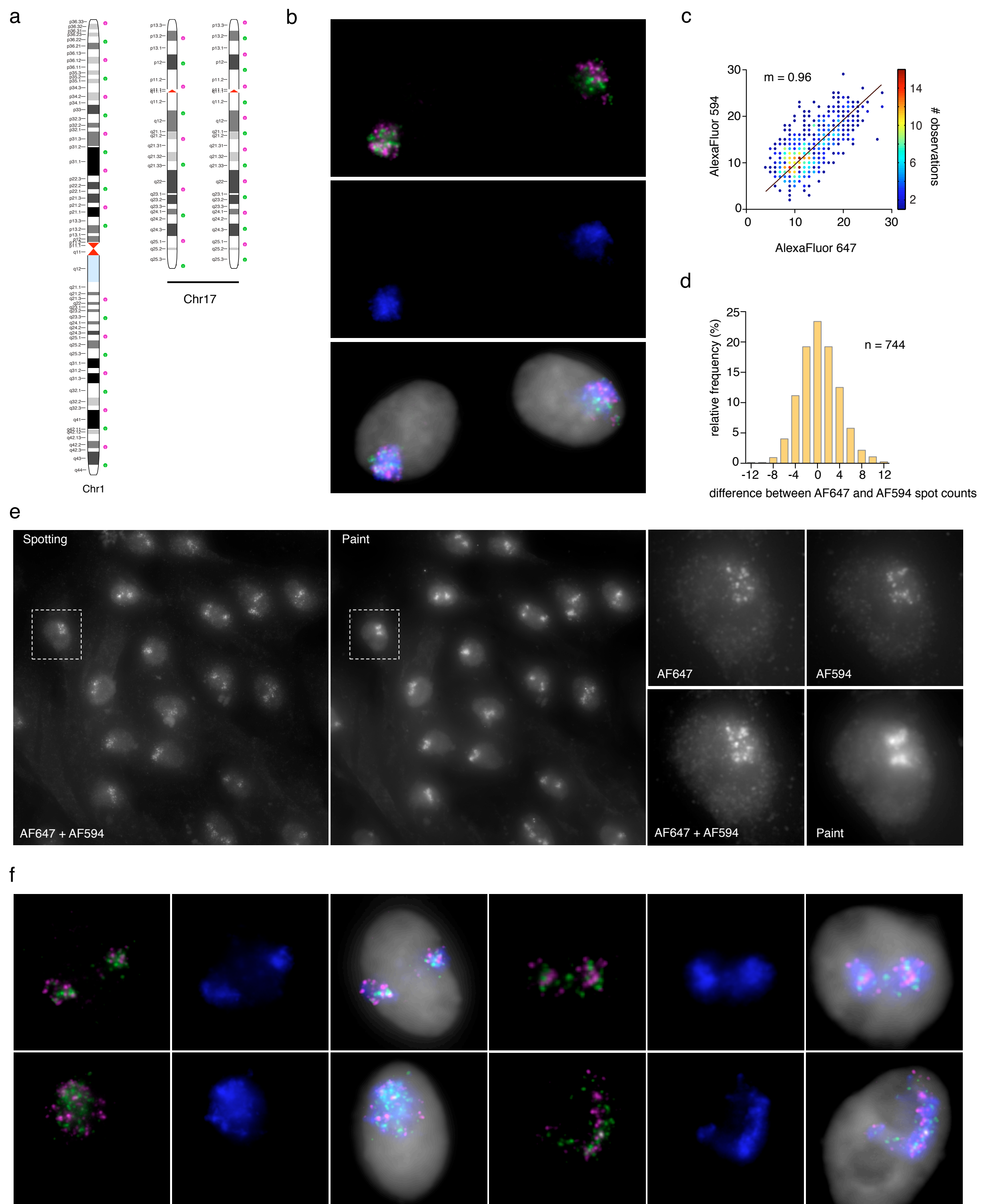
Supplementary Figure 1: Density of unique HD-FISH amplicons along the human (**a**) and mouse (**b**) genome. G-bands ideograms are displayed.



Supplementary Figure 2: HD-FISH cost analysis. (**a**) Cumulative cost as a function of the number of tests per probe in three different scenarios. Costs include synthesis of PCR primers and reagents for PCR, PCR purification, labeling, and preparation of hybridization solutions. Costs for other consumables (e.g. gloves, pipette tips, etc.) as well as for instrument depreciation and personnel are excluded. (**b**) Linear dependence of primer synthesis cost on the number of probes and amplicons per probe.

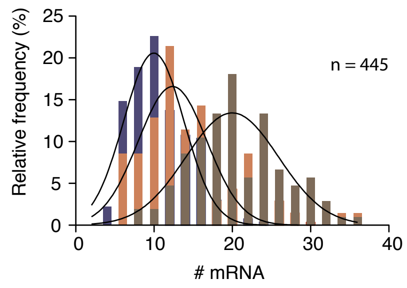


Supplementary Figure 3: Distribution of spot counts for HER2 probes of decreasing size in HME cells. The number of amplicons from which each probe was derived is shown. Numbers in brackets represents ETS values. n: number of cells analyzed.



Supplementary Figure 4: (**a**) Genomic location of individual probes forming the spotting probes for Chr1 and Chr17 described. Magenta: AlexaFluor647-labeled probes. Green: AlexaFluor594-labeled probes. G-bands ideograms and band names are displayed. (**b**) Chr1 spotting with twenty-two alternatively labeled HD-FISH probes (green and magenta), and simultaneous Chr1 painting (blue) in HME cells. (**c**) Inter-channel concordance for spot counts shown in **Figure 3b**. The linear regression fit line of slope m is shown. (**d**) Single-cell distribution of the difference between error estimates of HD-FISH spot counts measured

simultaneously in the AlexaFluor 594 and 647 channels. These data reflect gene copy number-invariant noise and demonstrate that the respective channel estimates are unbiased relative to each other (median value = 0) and symmetric (quartile skewness = 0). n: number of cells analyzed. (**e**) Representative field of view with HME cells simultaneously hybridized with a Chr17 spotting probe consisting of sixteen HD-FISH probes (left) and with a paint probe (right). Small panels: zoom-in view of the area marked by the white dashed square. AF647: AlexaFluor647; AF594: AlexaFluor594. (**f**) Comparison of Chr17 spotting (green and magenta) and painting (blue) in several Z-projections.



Supplementary Figure 5: Frequency distribution of HER2 mRNA counts in HME cells with 2 (purple), 3 (orange), or 4 (brown) HER2 loci detected by HD-FISH. Gaussian fits are superimposed onto histograms. n: number of cells analyzed.

SUPPLEMENTARY NOTE

In order for HD-FISH to attract a wide audience (including diagnostics laboratories and industrial parties), economic measures of its performance must be obtained in addition to scientific ones. For this purpose, we have performed cost effectiveness analysis in the following three scenarios. The first case is relevant to both research laboratories that occasionally need to perform DNA FISH on loci for which there are no ready-to-use probes commercially available, as well as to cytogenetics laboratories that wish to rapidly generate probes for rare rearrangements for which no commercial probes are available. The second case is relevant to diagnostics laboratories that perform a high volume of routine hybridizations against a limited number of loci, and that wish to reduce reagents costs. The third case is relevant to policy makers and healthcare systems organizers that are willing to financially support initiatives aiming at improving the efficiency of high quality diagnostics by cutting reagents costs and optimizing processes.

CASE 1. Let us imagine the case in which a particular locus (or a small number of loci) for which no commercial probes are available needs to be analyzed for research or diagnostic purposes. An investment of approx. $1,880 allows purchase of primers for 24 amplicons (which allows robust signal quantification as shown in the new **Supplementary Fig. 2a**) and for all the reagents needed for PCR, labeling, purification, and hybridization. This investment enables approx. 900 tests (i.e. hybridizations with 20 ng of probe in 20 μL buffer) at no extra costs, after which the cumulative cost grows in a pseudo-linear, step-wise manner at very slow pace, mostly driven by the need to periodically replenish PCR and labeling reagents (**Supplementary Fig. 2a**). An extra starting investment of $900, for example, enables ten HD-FISH probes to be simultaneously prepared in the same cost-effective manner as outlined above, allowing 90 tests per probe to be performed at no extra cost (**Supplementary Fig. 2a**). In comparison, custom design and BACs-based synthesis of ready-to-use probes for the same number of loci enabling at least 90 tests per probe would at minimum require $8,000-10,000 (data based on quote requests to several DNA FISH probes manufacturers, including Empire Genomics, Kreatech Diagnostics, and Abnova).

CASE 2. Let us imagine the case in which a limited set of probes (e.g. 20) is routinely used for diagnostic purposes. An investment of less than $5,800 allows purchasing primers sets for 20 probes, each consisting of 48 amplicons (which based on our results allows robust signal quantification in cells as well as tissues), as well as reagents needed for PCR, labeling,

purification, and hybridization. After 45 initial tests per probe at no extra cost, the cumulative cost rises in a pseudo-linear, step-wise manner, mostly driven by the need to periodically replenish PCR and labeling reagents (**Supplementary Fig. 2a**). The cost effectiveness of this approach is obvious by examining the cumulative cost (approx. $8,000) after 200 tests per probe (i.e. 4,000 tests in total, assuming that each probe is used with the same frequency), a figure perfectly within the range of average annual volume of tests performed by medium-sized cytogenetics laboratories. In contrast, performing the same volume of tests for 20 different ready-to-use probes periodically purchased from existing commercial providers would require a budget at least 5-fold higher (as an example, $2,500 is the price charged by a top manufacturer for a ready-to-use probe for 50 tests. Even assuming a 20% discount from the price list for large order volumes, purchase of probes for 20 different loci for 200 tests/probe would require $160,000).

CASE 3. Though surely cost effective in the cases depicted above, the start-up cost of HD-FISH linearly grows with the number of primers (i.e. probes), quickly exceeding the financial capabilities of a single laboratory (**Supplementary Fig. 2b**). Moreover, even at the lowest possible synthesis scale, a large fraction of primers would remain unused, as it is basically impossible for a single laboratory to reach the maximum number of tests (> 8 millions hybridizations using 20 ng of probe in 20 μL buffer per test) that can be performed from a 96-well plate containing 12 nmol primers per well. Instead, a consortium funded by public resources and administered as a non-profit repository could build up a primers stock covering the human and mouse genomes for less than $20 million. This budget is well within the average funding that public agencies have been granting to such type of initiatives in the U.S. as well as in the E.U. This would enable laboratories worldwide to purchase primers or even pre-made PCR reactions at low price, therefore greatly facilitating high-quality FISH-based research and diagnostics services, especially in emerging economies.