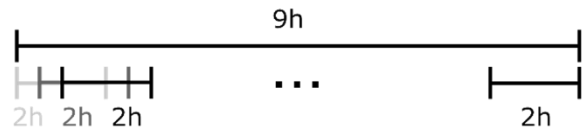
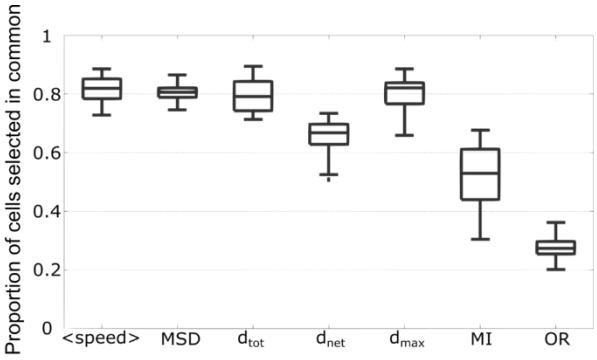
a



b



c

Motility-related parameters	Definition
Distance between track points	$d(\vec{x}_i, \vec{x}_j) = \sqrt{\sum_{d=1}^{D}(x_{i,d} - x_{j,d})^2}$ where $D = 2$ *or* 3
Average speed	$< speed > = \frac{1}{N}\sum_{i=1}^{N}\frac{d(\vec{x}_i, \vec{x}_{i+1})}{dt}$
Total distance	$d_{tot} = \sum_{i=1}^{N-1} d(\vec{x}_i, \vec{x}_{i+1})$
Net distance	$d_{net} = d(\vec{x}_1, \vec{x}_N)$
Max distance	$d_{max} = \max_{i,j} d(\vec{x}_i, \vec{x}_j)$
Mean-square displacement	$MSD = \frac{1}{N-1}\sum_{i=1}^{N-1} d(\vec{x}_i, \vec{x}_{i+1})^2$
Meandering index	$MI = \frac{d_{net}}{d_{tot}}$
Outreach ratio	$OR = \frac{d_{max}}{d_{tot}}$

Fig. S1. Reduction of the tracking experiment time and selection of a motility parameter. (a) Schematic of the experiment. A 9h time-lapse acquisition of migrating cells was performed. Cells were tracked and classified according to different motility-related parameters. Cells were also classified over widows of 2h included in the full acquisition. **(b)** Proportion of cells identified as motile for both a 9h-long time-lapse and all the 2h-hour long videos for the different parameters. The parameter that offered the greatest proportion of common selected cells was *dmax*. **(c)** list and mathematical definition of the different motility-related parameters.

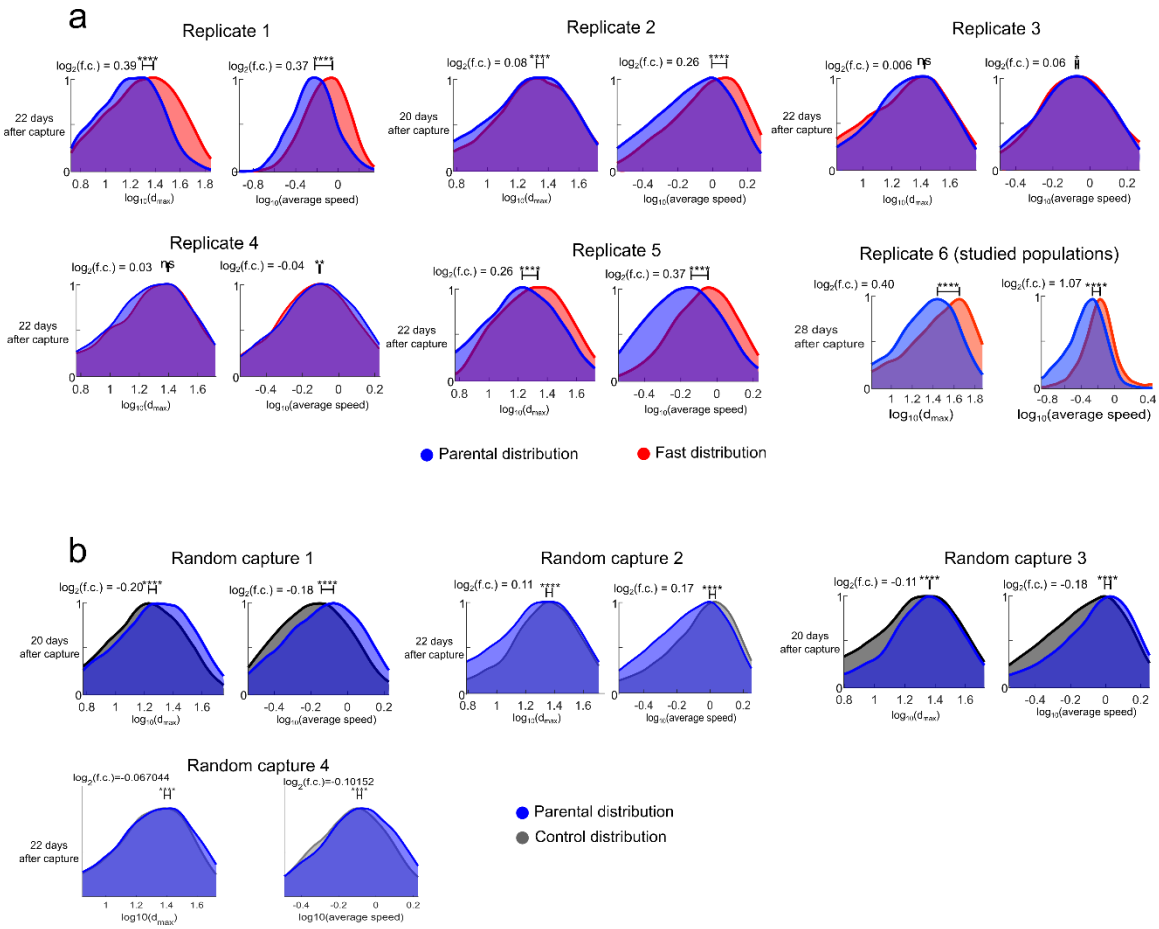


Fig. S2. Capture replication and random captures (a) Replicates of fast captured cells. The 6th replicate is the one used for the main study. Y-axis is the probability density function normalized by the mode. (b) Random selection replicates. The captured cells represent 5% of the total population. We captured faster cells only in the second random capture. Y-axis is the probability density function normalized by the mode.

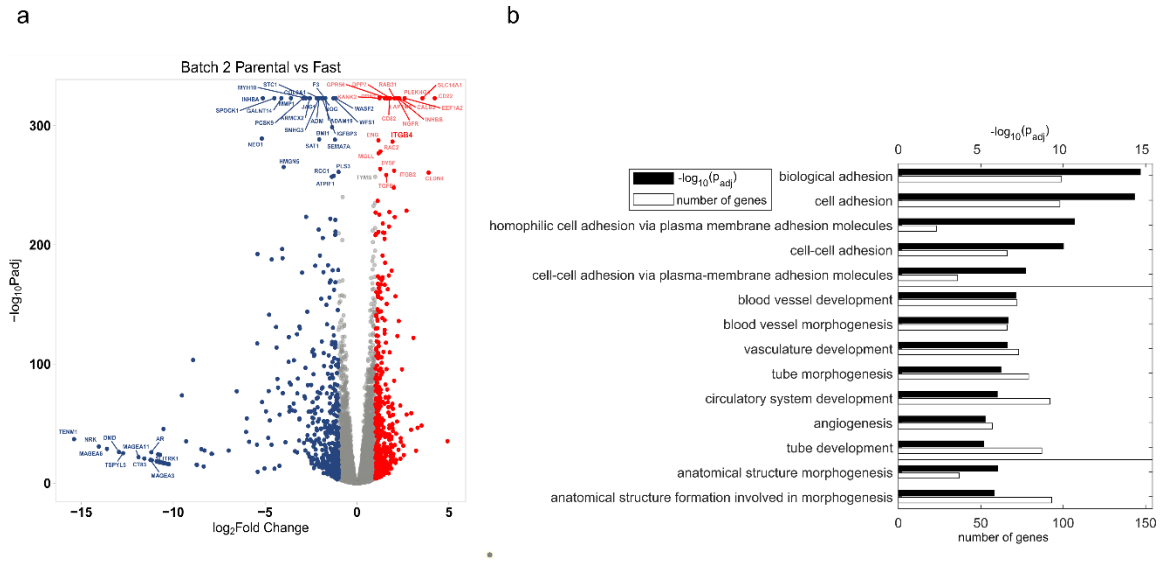


Fig. S3. (a) Volcano plot over the transcripts highlighting the most significative genes from the second RNS sequencing batch. (b) Groups of the most important GO biological process terms from the second analysis highlights terms related to adhesion.

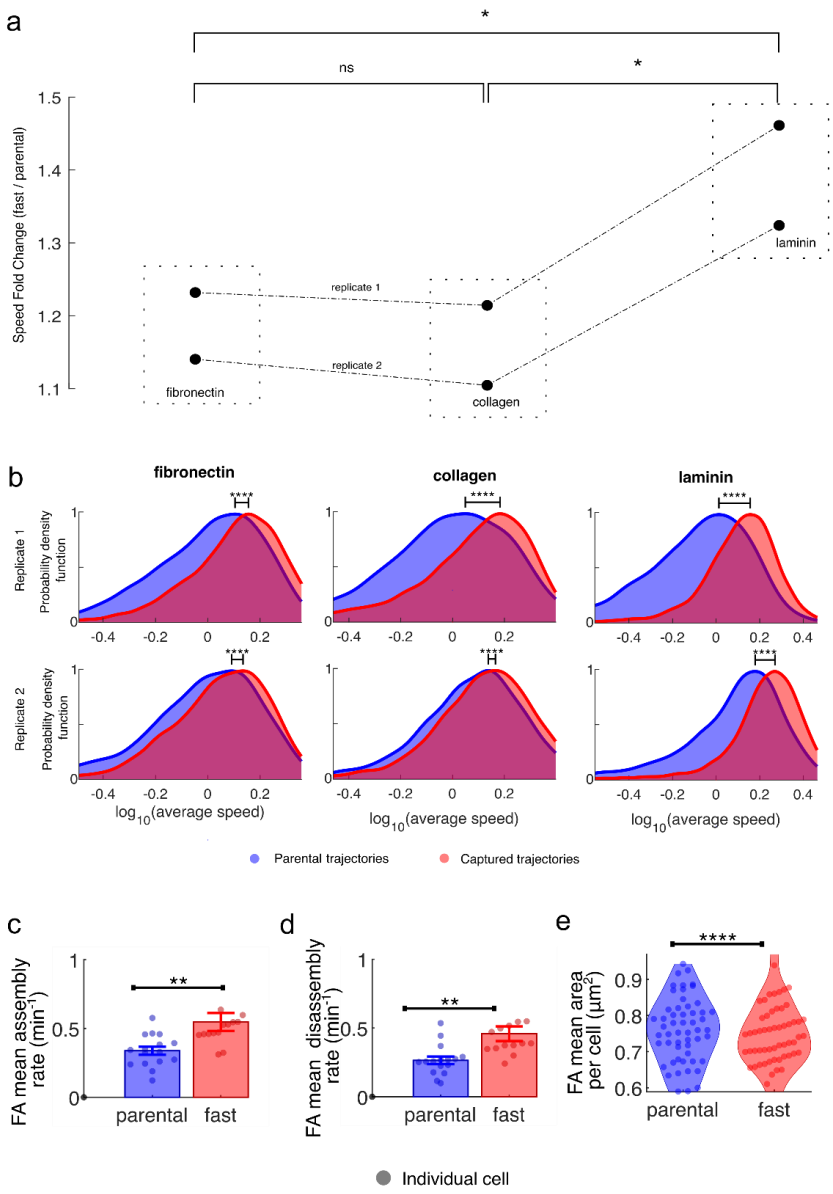
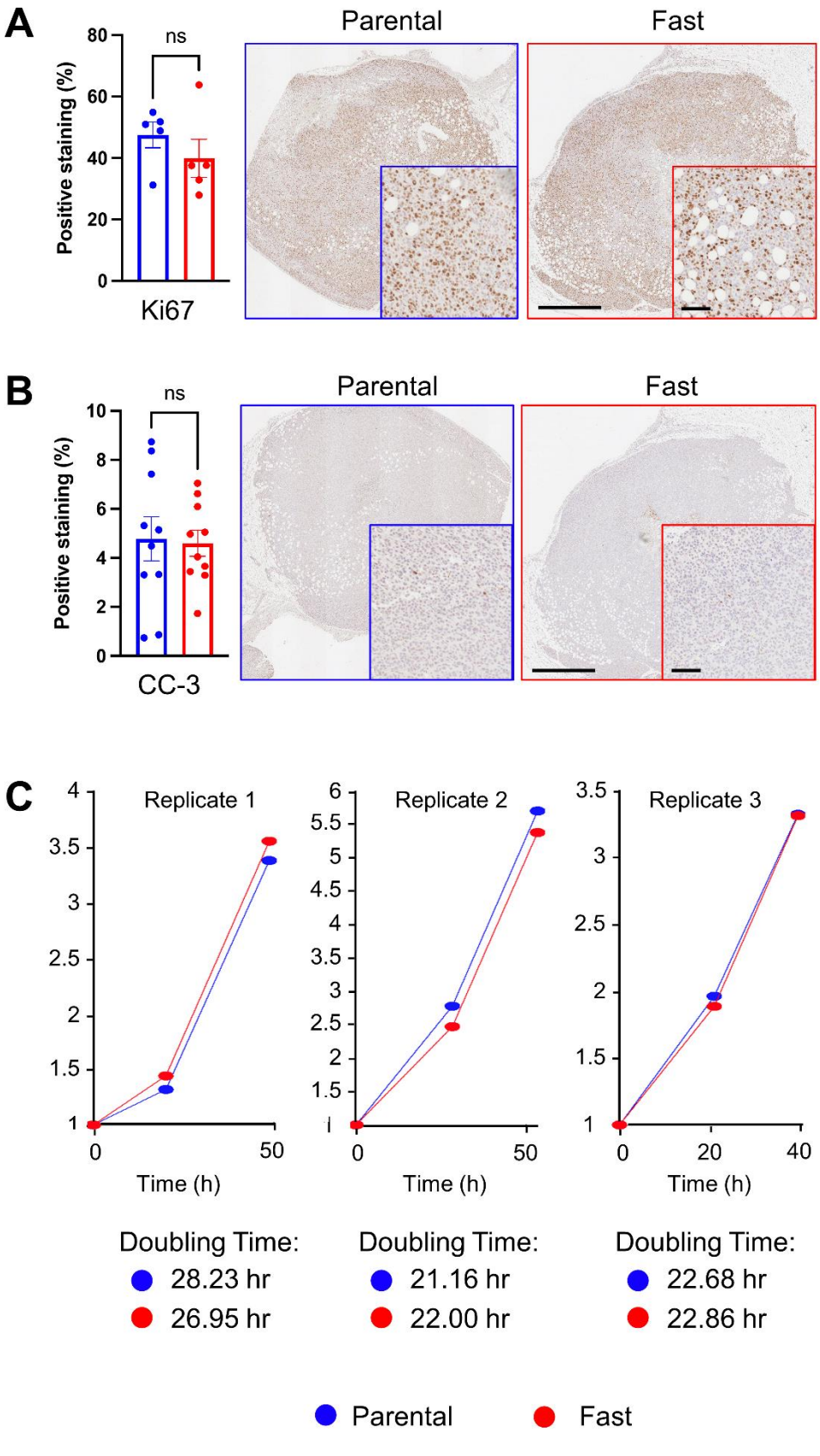
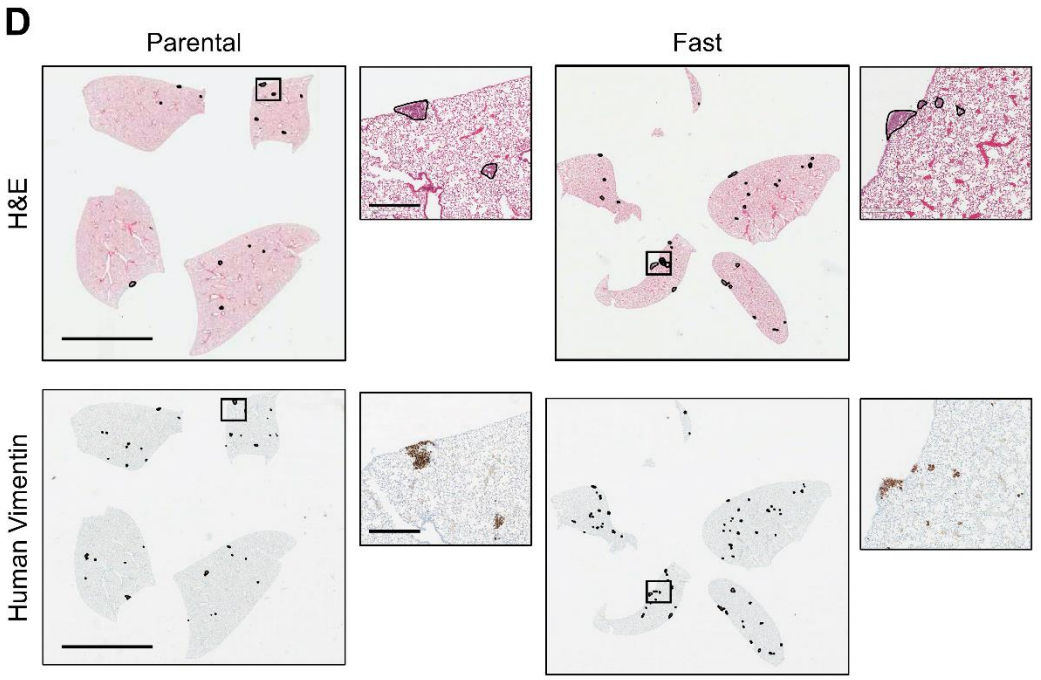


Fig. S4. Tracking experiments on three ECM coatings: fibronectin, collagen and laminin-5. (a) Fold change in average speed between parental cells and fast cells for all the coatings and two replicates. The difference in speed between the fast and parental cells is significatively larger for laminin compared to collagen ($p = 0.02$) and fibronectin ($p = 0.03$). The difference is not significative between fibronectin and collagen ($p = 0.1$). (b) Average speed distributions for the indicated populations and coating for the two replicates. Assembly (c) and disassembly (d) rates for cell populations on laminin-332. (e) FA size averaged per cell on laminin- 332. Only one of the triplicates is shown here for simplicity, similar results were obtained in all replicates.





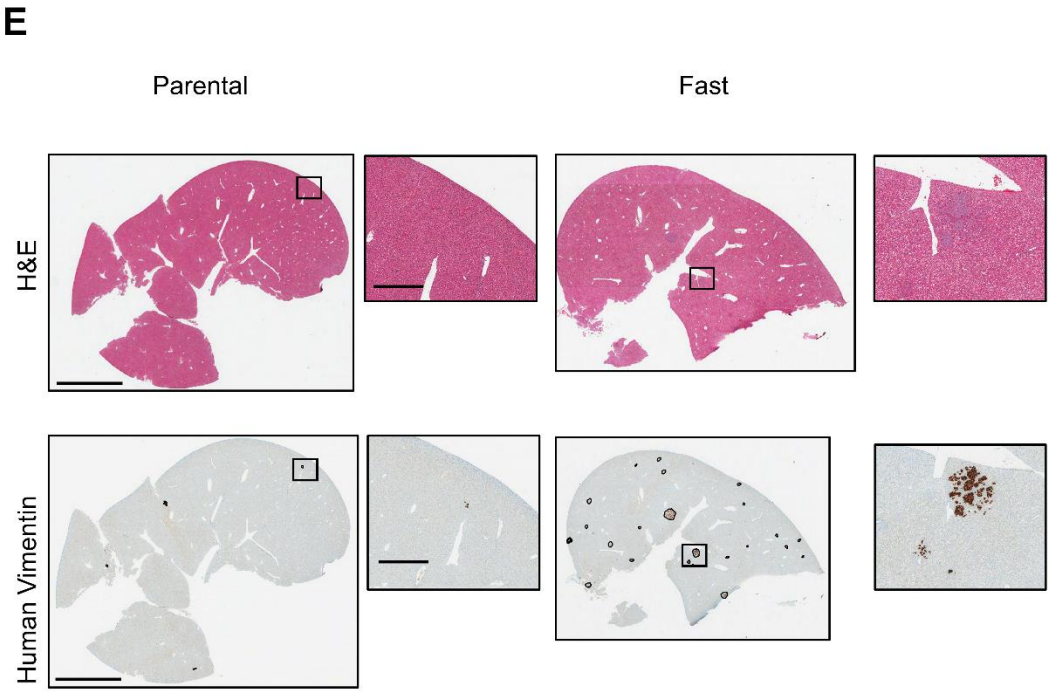
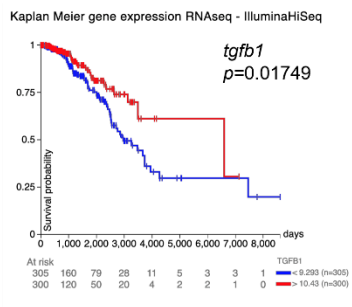
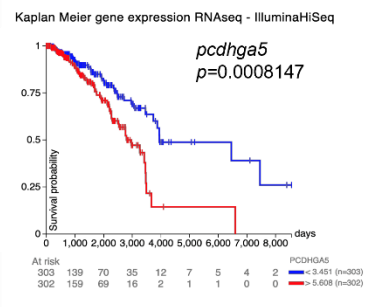
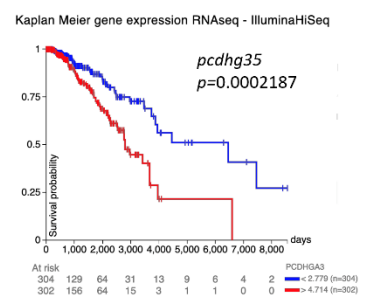
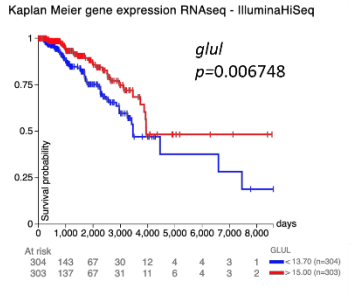


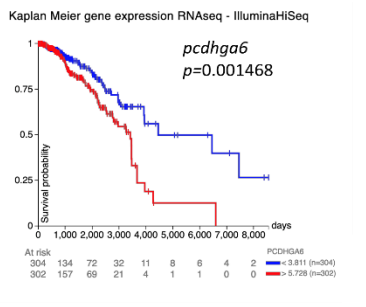
Fig. S5. At endpoint, tumor tissue was harvested from mice bearing mammary tumors derived from parental and fast-selected MDA-MB-231 cells. Immunohistochemistry was performed using Ki67 and Cleaved Caspase 3 (CC-3) antibodies to monitor proliferation (A) and apoptosis (B), respectively. Scale bar represents 1mm in the low magnification image and 100um in the high magnification image. (C) *In vitro* growth curves of parental and fast MDA-MB-231 cells seeded on plastic dishes. Doubling times for each population within each replicate are shown. To calculate doubling times, the curves were fitted with the equation: $n(t)/n0 = 2t/Td$, where n(t) is the number of cells in the dish at time t, n0 is the initial number of cells, t is the time in hours, and Td is the doubling time in hours. P = 0.98 was obtained by comparing the calculated doubling times for each parental and fast MDA-MB-231 replicate with a t-test. (D-E) Representative images of lungs (D) or liver (E) isolated from mice bearing fast- and parental-derived MDA-MB-231 tumors at end-point. The tissues were stained with H&E or Vimentin to identify metastatic lesions. The scale bar in the low magnification image represents 6mm and 400 µm in the higher magnification image.

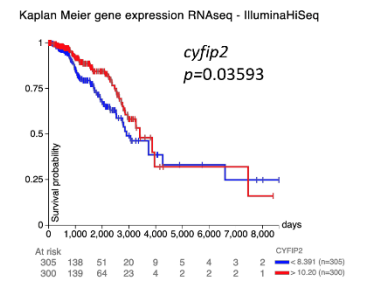


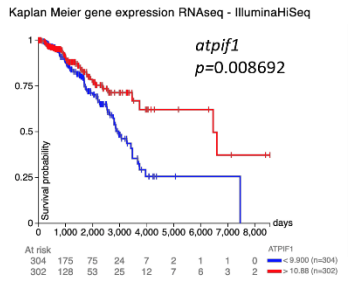


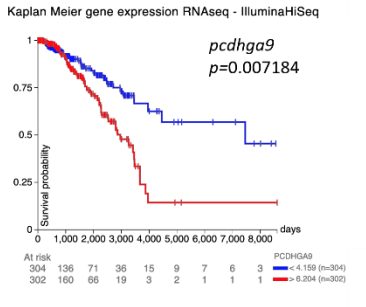


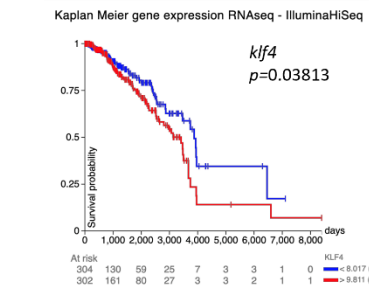


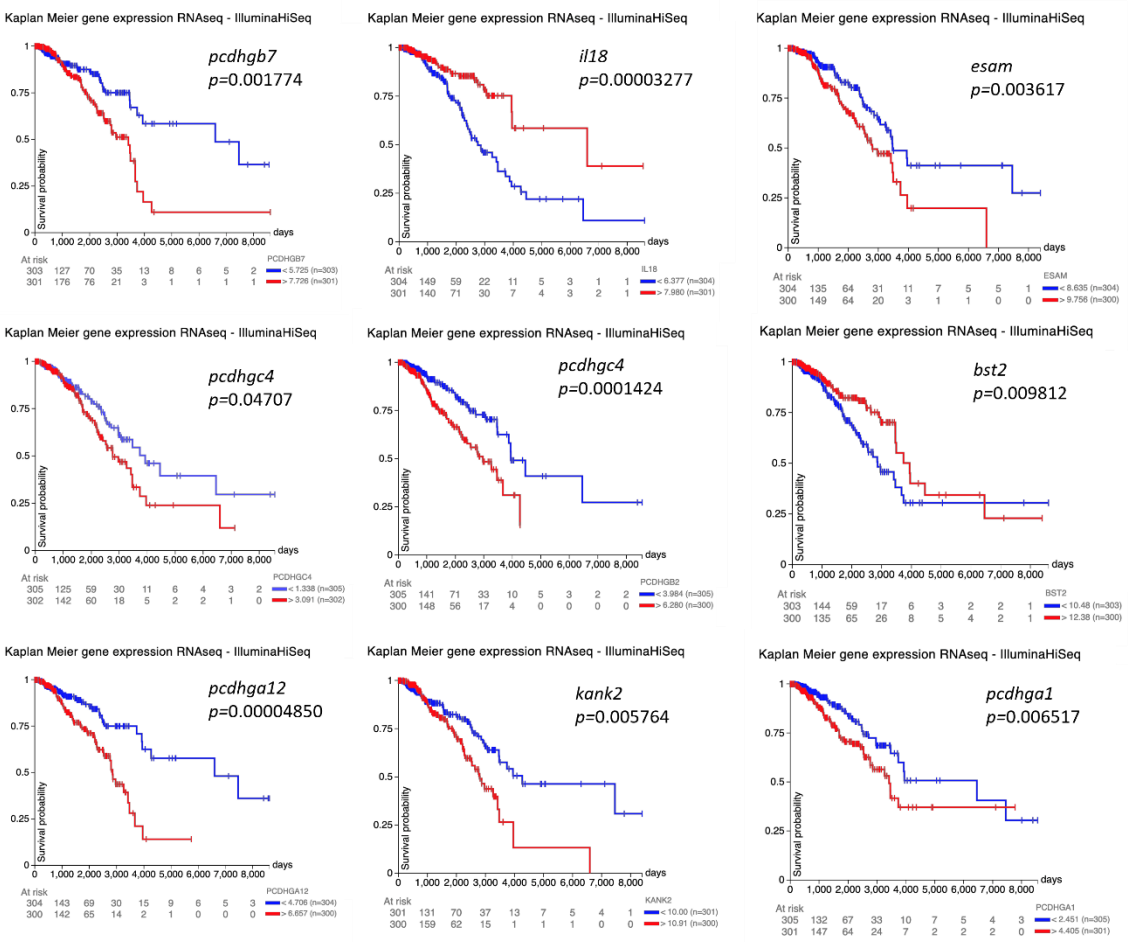











Kaplan Meier gene expression RNAseq - IlluminaHiSeq
pcdhgb7
p=0.001774
Survival probability
days
At risk
PCDHGB7
303 127 70 35 13 8 6 5 2
301 176 76 21 3 1 1 1 1
< 5.725 (n=303)
> 7.726 (n=301)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
il18
p=0.00003277
Survival probability
days
At risk
IL18
304 149 59 22 11 5 3 1 1
301 140 71 30 7 4 3 2 1
< 6.377 (n=304)
> 7.980 (n=301)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
esam
p=0.003617
Survival probability
days
At risk
ESAM
304 135 64 31 11 7 5 5 1
300 149 64 20 3 1 1 0 0
< 8.635 (n=304)
> 9.756 (n=300)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
pcdhgc4
p=0.04707
Survival probability
days
At risk
PCDHGC4
305 125 59 30 11 6 4 3 2
302 142 60 18 5 2 2 1 0
< 1.338 (n=305)
> 3.091 (n=302)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
pcdhgc4
p=0.0001424
Survival probability
days
At risk
PCDHGB2
305 141 71 33 10 5 3 2 2
300 148 56 17 4 0 0 0 0
< 3.984 (n=305)
> 6.280 (n=300)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
bst2
p=0.009812
Survival probability
days
At risk
BST2
303 144 59 17 6 3 2 2 1
300 135 65 26 8 5 4 2 1
< 10.48 (n=303)
> 12.38 (n=300)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
pcdhga12
p=0.00004850
Survival probability
days
At risk
PCDHGA12
304 143 69 30 15 9 6 5 3
300 142 65 14 2 1 0 0 0
< 4.706 (n=304)
> 6.657 (n=300)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
kank2
p=0.005764
Survival probability
days
At risk
KANK2
301 131 70 37 13 7 5 4 1
300 159 62 15 1 1 1 0 0
< 10.00 (n=301)
> 10.91 (n=300)
Kaplan Meier gene expression RNAseq - IlluminaHiSeq
pcdhga1
p=0.006517
Survival probability
days
At risk
PCDHGA1
305 132 67 33 10 7 5 4 3
301 147 64 24 7 2 2 2 0
< 2.451 (n=305)
> 4.405 (n=301)

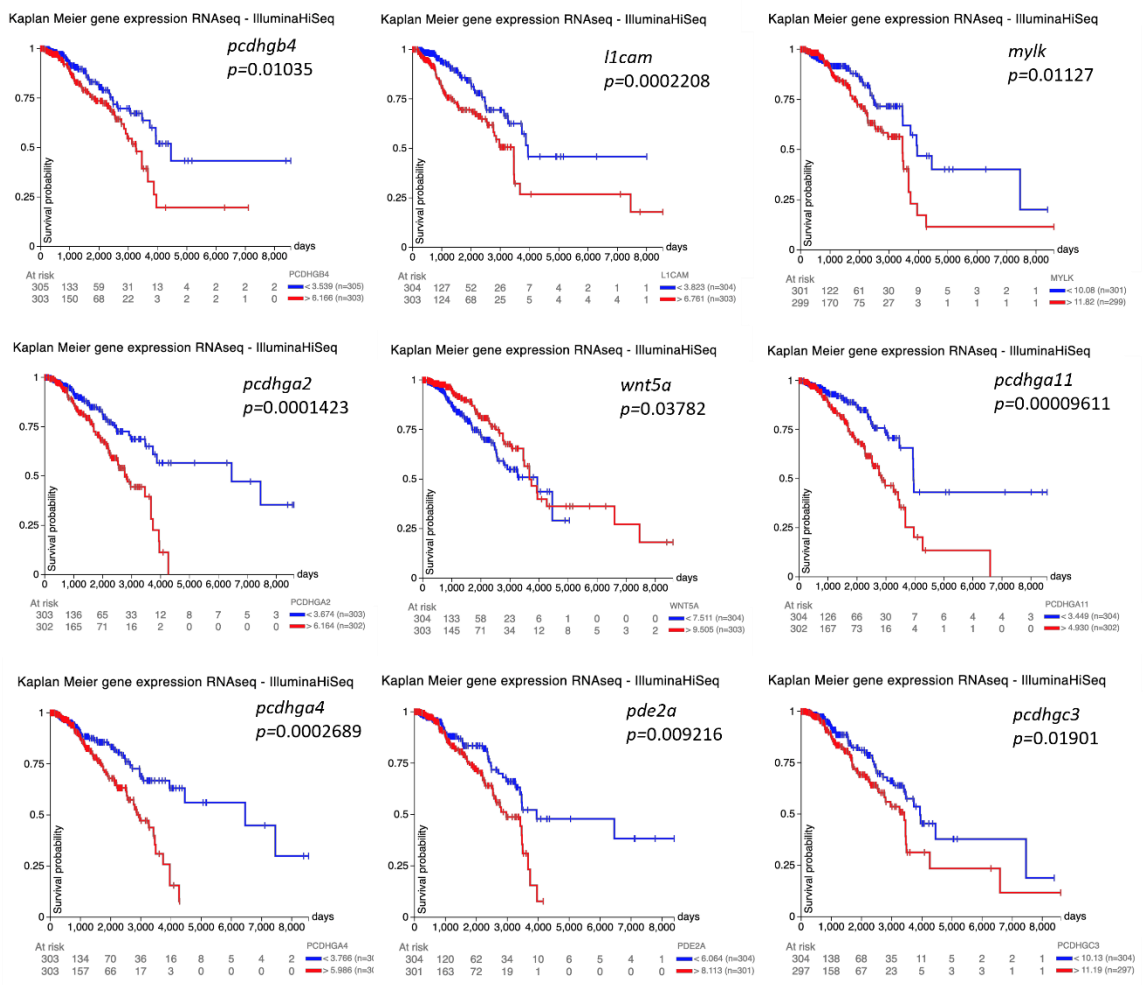


Fig. S6. Kaplan-Meier curves showing overall survival curves associated with the expression of the indicated genes in the TCGA dataset.

Table S1. Differential expression of fast versus parental samples. Positive log2 fold change means that there is more expression in the "fast" group.

Click here to download Table S1

Table S2. List of antibodies used in this study

Click here to download Table S2