Critical roles for EGFR and EGFR-HER2 clusters in EGF binding of SW620 human carcinoma cells

Authors
Adam J. M. Wollman[1,2]†, Charlotte Fournier[3,4,] †, Isabel Llorente-Garcia[5,] †, Oliver Harriman[3,] †, Alex L. Payne-Dwyer[1], Sviatlana Shashkova[1,6], Peng Zhou[7], Ta-Chun Liu[8], Djamila Ouaret[8], Jenny Wilding[8], Akihiro Kusumi[7], Walter Bodmer[8] and Mark C. Leake[1,9,*]

Affiliations
[1] Department of Physics, University of York, York, United Kingdom.
[2] Current address: Biosciences Institute, Newcastle University
NE2 4HHO, United Kingdom.
[3] Department of Physics, Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, United Kingdom.
[4] Okinawa Institute of Science and Technology Graduate University (OIST), 1919 Tancha, Onna-son, Okinawa, Japan 904-0495.
[5] Current address: Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom
[6] Current address: Department of Physics, University of Gothenburg, 412 96 Gothenburg, Sweden.
[7] Membrane Cooperativity Unit, OIST, 1919 Tancha, Onna-son, Okinawa, Japan 904-0495.
[8] MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom.
[9] Department of Biology, University of York, York, United Kingdom.
† These authors contributed jointly to this work
[*] Correspondence should be addressed to M.C.L: mark.leake@york.ac.uk

Abstract
Epidermal growth factor (EGF) signaling regulates normal epithelial and other cell growth, with EGF receptor (EGFR) overexpression reported in many cancers. However, the role of EGFR clusters in cancer and their dependence on EGF binding is unclear. We present novel single-molecule TIRF microscopy of: (i) EGF and EGFR in living cancer cells, (ii) the action of anti-cancer drugs that separately target EGFR and HER2 on these cells, and (iii) EGFR-HER2 interactions. We selected human epithelial SW620 carcinoma cells for their low level of native EGFR expression, for stable transfection with fluorescent protein labeled EGFR, and imaged these using single-molecule localization microscopy to quantify receptor architectures and dynamics upon EGF binding. Prior to EGF binding we observe pre-formed EGFR clusters. Unexpectedly, clusters likely contain both EGFR and HER2, consistent with co-diffusion of EGFR and HER2 observed in a different model CHO-K1 cell line, whose stoichiometry increases following EGF binding. We observe a mean EGFR:EGF stoichiometry ~4:1 for plasma-membrane-colocalized EGFR-EGF that we can explain using novel time-dependent kinetics modeling, indicating preferential ligand binding to monomers. Our results may inform future cancer-drug developments.

Keywords
Single-molecule, cancer, super-resolution, receptors, inhibitors

Introduction
Epidermal growth factor receptor (EGFR) is a cell surface receptor essential for cell growth and differentiation, with its disregulation implicated in several carcinomas (1), hence a target for numerous cancer drugs. Human EGFR or ERBB1, ('ErB1'or 'HER1') is a protein of the receptor

tyrosine kinase (RTK) family and the ERBB subfamily with three other ERBB members, ERBB2 ('ErbB2' or 'HER2'), ERBB3 ('ErbB3' or 'HER3') and ERBB4 ('ErbB4' or 'HER4'), expressed in the plasma membranes of mainly epithelial cells (2). EGFR has an extracellular region with subdomains I-IV, of which I and III participate in ligand binding (3). The extracellular region is connected to a cytoplasmic domain containing a tyrosine kinase.

There are 11 ligands that can bind to ERBB proteins, including epidermal growth factor (EGF) which binds to EGFR (4). Ligand binding induces receptor dimerization/clustering, resulting in activation following tyrosine residue autophosphorylation that initiates signaling reactions to stimulate cell growth, differentiation and proliferation. Structural evidence indicates that activation is preceded by EGF binding to EGFR monomers that induces a conformational change by removing interactions that autoinhibit EGFR dimerization (5). Binding studies of full-length receptors suggest negative cooperativity, mediated through an intracellular juxta-membrane domain (6), as do radioligand-binding and phosphorylation assays (7, 8).

An early single-molecule fluorescence imaging study using model human epidermoid cell line A431 published in 2000 reported binding of single EGF to a pre-formed EGFR dimer, followed by a second molecule to form a 2:2 complex (9), however later findings from *Xenopus* oocytes suggested in that system that the majority of EGFR was present as a monomer (10). Other studies have instead reported observations of pre-formed EGFR oligomers using a range of methods comprising antibody-labeled EGF (11), Förster resonance energy transfer (FRET) (12), autocorrelation analysis (13), bimolecular fluorescence complementation (BiFC) (14), pixel brightness analysis (15) and single-molecule live cell light microscopy (16, 17). The clustering and oligomeric states of EGFR are also complex since they may involve cooperativity not only between EGFR but also other ERBB proteins (14). EGFR's clustering state before and after EGF binding under physiological conditions has remained contentious due to limitations in simultaneous data on stoichiometries of interacting receptors and ligands, to a dependence of EGF expression on EGFR clustering, to the common simultaneous presence of fluorescently labelled EGFR and dark EGFR, and to the existence of species-specific cell-line differences.

Other ERBB receptors such as HER2 have been detected in monomeric, dimeric and higher-order clusters in human breast cancer cells (18), and in clusters of 2-4 HER2 molecules in fixed breast cancer cell lines determined using super-resolution fluorescence microscopy (19). Furthermore, several light microscopy studies have suggested interactions of EGFR with other ERBB receptors in human cancer cells. For example, EGFR and HER2 co-express in human bladder cancer and colorectal cancer cell lines (20, 21) and human breast cancer cells SKBR3. In SKBR3, EGFR and HER2 expression levels can jointly increase in large membrane protrusions (22), hinting at the possibility of EGFR-HER2 heterodimers. EGFR-HER2 interactions inside lipid rafts in SKBR3 cells have also been proposed (23). The presence of pre-formed homo- and heterodimers of different ERBB family members, including EGFR and HER2, has also been inferred from lysate analysis of transfected CHO cells (24). Interactions of EGFR with the hepatocyte growth factor receptor HGFR (a.k.a. MET) have also been inferred from single-molecule imaging where increased colocalization and decreased diffusion was observed in live HeLa and BT-20 cells after EGF stimulation (25).

Here, we used two-colour single-molecule total internal reflection fluorescence (TIRF) microscopy for super-resolved single-molecule localization microscopy (SMLM) on live human colon carcinoma cells stably expressing EGFR-GFP in the presence of tetramethylrhodamine (TMR) conjugated to EGF (Figure 1). Supported by predictions from Monte Carlo simulations, we find that prior to EGF binding, EGFR forms clusters with a modal stoichiometry of 6 molecules but extending to 10s of molecules, adding to an emerging consensus that pre-formed EGFR clusters exist prior to EGF activation. Following EGF binding, we see clusters with a 3-fold higher stoichiometry. We find that EGF-bound EGFR clusters have a relative stoichiometry ratio for EGFR:EGF of approximately 4:1, which we interpret using a new time-dependent kinetics model that shows preferential ligand binding to receptor monomers with no binding to dimers. We present the first single-molecule light microscopy observations of the effect on live

human cancer cells of anti-cancer immunotherapy drugs cetuximab (26) and trastuzumab (27) which specifically and separately inhibit EGFR activation by targeting either EGFR or HER2 respectively. We find that both promote an increase in EGFR cluster stoichiometry and a decrease in diffusion coefficient after addition of EGF. Compared to untreated cells, treatment with both either drug in addition to EGF results in increased numbers of EGFR molecules in a cluster and in a higher diffusion coefficient for EGF-bound EGFR clusters, which may reflect cluster compaction. Additionally, we present novel dual-colour single-molecule TIRF imaging of EGFR-HER2 interactions from live CHO-K1 cells that contain both fluorescently labeled EGFR and HER2. These data show that EGFR and HER2 in a model cell line interact transiently before EGF binding with a dwell time of several hundred milliseconds. Taken together, these observations show that EGFR clusters comprise a mixture of EGFR and HER2, to be compared with indirect findings of heterodimer formation in SKBR3 breast cancer cells from correlative fluorescence microscopy and liquid phase electron microscopy (22). Our results provide new insights into architectures, dynamics and interactions of EGFR molecules overexpressed in carcinoma cells. Instead of a simplified picture for EGFR function in terms of monomer and dimer states, they indicate higher levels of complexity which hitherto has not been addressed explicitly. Given the nature of the EGFR pathway as an anti-cancer drug target, our results may inform the development of new therapeutic strategies to treat cancer.

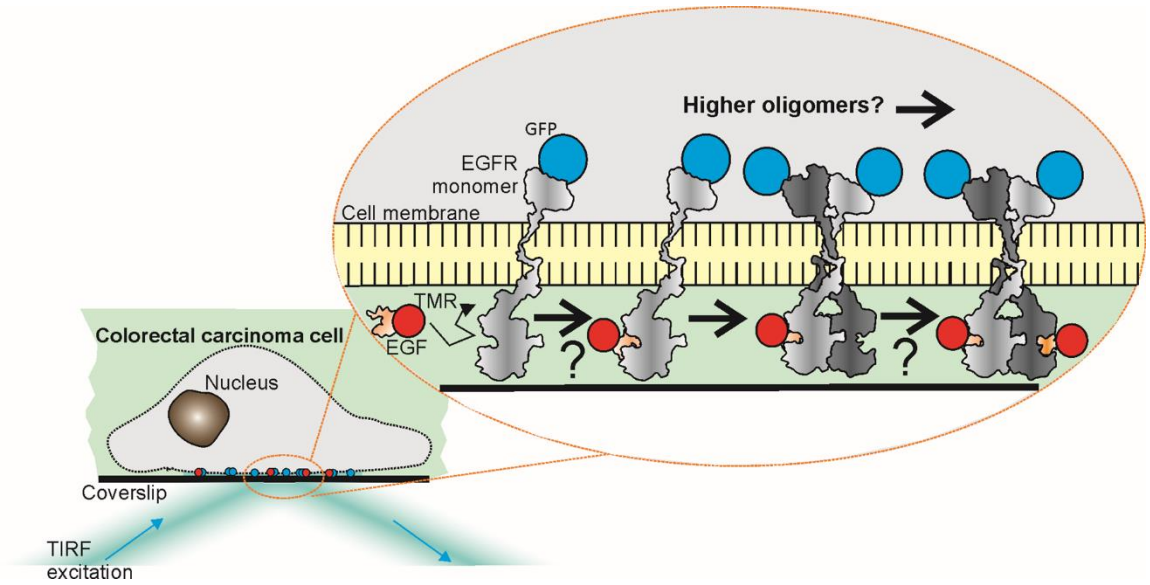


Figure 1. Visualizing EGF-EGFR in SW620 cells. *Current models to explain EGFR activation encompass different binding rates of EGF to EGFR monomers and dimers, and binding cooperativity between EGF and EGFR. However, questions remain as to the role of EGFR clusters in cancer cells and their dependence on EGF binding. Here we used TIRF microscopy of GFP-labelled EGFR (blue) and TMR-labelled EGF (red) to enable SMLM to address these questions.*

Results

Single-molecule microscopy reveals EGFR clusters before EGF binds in SW620 cells.

To visualize EGFR molecules in live cancer cells, we generated a human colon carcinoma cell line stably expressing EGFR-GFP. Immortalized cell line SW620, deriving from a human lymph node metastasis from an adenocarcinoma of the colon (28), was selected from a colon carcinoma library (29) for its very low EGFR expression (Figure S1 and S2) consistent with previous recent findings (30–33) and low expression of the most common EGFR ligands, including TGFA. TGFA has been reported to be expressed in SW620 cells from one study published in 1987 (34),

however our recent high-precision microarray measurements indicate only very low levels (Supplementary details our SW620 microarray results for all EGFR ligands). EGFR-GFP kinase activity in SW620 cells was confirmed by observing increased phosphorylation of EGFR downstream targets, ERK1/2, in response to EGF (Figure S2b).

We optimized a home-built TIRF microscope (Figure S4) for single-molecule detection, confirmed using an *in vitro* surface assay (35) in which GFP was antibody-conjugated to a glass coverslip (Figure S5). After approximately 1 s of laser illumination, bright spots (fluorescent foci) on our image sequences exhibited step-wise photobleaching (Figure S5) indicating the presence of single GFP molecules. Single fluorophore brightness values were quantified by analyzing distributions of fluorescent foci intensity values (Figure S5c).

We applied our optimized TIRF microscopy to transfected SW620 cells in serum-free medium without addition of EGF. We observed fluorescent foci at a surface density of 0.1-0.4 per μm^2 in the basal plasma membrane in contact with the glass coverslip (Figure 2a and Figure S6) with a mean of 66±28 (s.d.) foci per cell. In most cells, foci could be detected across the full extent of the basal membrane and exhibited a smooth surface topography consistent with earlier scanning electron microscopy performed on SW620 cells (36). We tracked foci over several seconds to approximately 40 nm spatial precision using home-written tracking software (37) (movie S1).

Foci image widths were on average within 10% of those observed for single GFP *in vitro* (~250 nm half width at half maximum). However, their brightness was greater than that expected for monomeric GFP, with fluorescence intensity traces exhibiting multiple stochastic photobleaching steps (Figure 2b) indicative of several molecules within each EGFR cluster. We could determine the stoichiometry of these foci by dividing their initial brightness by that of a single GFP (35). The mean brightness of a single GFP was measured *in vivo* by quantifying the foci brightness towards the end of a photobleach trace, when only one photoactive molecule remained. The *in vivo* single-GFP brightness obtained in this way was within 15% of that measured *in vitro*, confirming accurate single-molecule detection *in vivo* (Figure S5c).

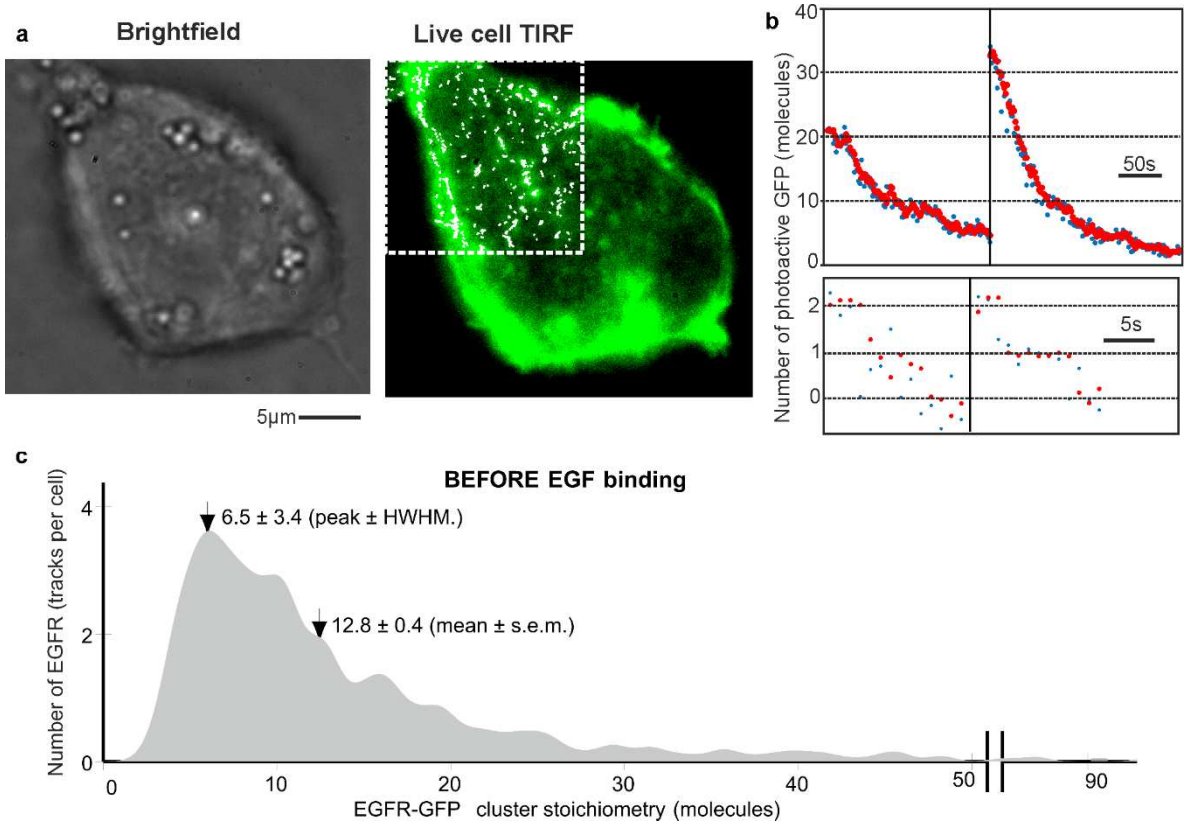


Figure 2. EGFR cluster stoichiometry in SW620 cells before EGF binding. ***(a)*** *Transfected SW620 cell showing GFP (green) and overlaid tracking (white) on top left corner.* ***(b)*** *Photobleaching intensity traces from tracked EGFR-GFP clusters with stoichiometries of several tens of molecules (upper), down to two molecules (lower).* ***(c)*** *Distribution of EGFR cluster stoichiometry before EGF binding showing peak at approximately 6 molecules and a mean of 12.8 molecules, with N=19 cells, and 1,250 cluster tracks in total (66 tracks per cell), corresponding to approximately 850 tracked EGFR per cell on average.*

By integrating total fluorescence GFP intensity in each cell and correcting for native autofluorescence we estimate the total copy number is approximately 200,000 EGFR-GFP molecules per cell. Tracked foci brightness values indicated that they comprise clusters of EGFR with a broad stoichiometry distribution, across different cells and within the same cell, with a range 2-90 EGFR molecules per cluster, with a peak value of approximately 6 and a mean of 12.8±0.4 molecules (±s.e.m) (Figure 2c). We did not detect any monomeric EGFR-GFP before adding EGF from >1,000 tracks in 19 different cells (Table S1), despite our microscope having single GFP sensitivity *in vivo* and *in vitro* under the same imaging conditions (Figure S5). We considered whether the absence of detected monomers and broad range of stoichiometry could be due to the random optical overlap of lower stoichiometry EGFR clusters in our diffraction-limited images. We modeled this effect by convolving a Poisson distribution calculated from the overlap probability (38) with the brightness distribution of a cluster in a range of different stoichiometry states (similar to earlier studies (17, 39)). The simulated EGFR cluster stoichiometry distributions due to optical overlap for 1-4 molecules per cluster had a poor resemblance to the experimental stoichiometry distribution (Figure S7). However, simulating a cluster stoichiometry randomly sampled from a second Poisson distribution with peak value equal to 4 molecules per cluster, but extending to tens of molecules per cluster, resulted in reasonable predictions which could account for approximately 90% of the observed variance (R^2=0.88) in the experimental stoichiometry distribution (Figure S7). This suggests that many EGFR foci are formed from clusters with a broad stoichiometry distribution. Rather than EGFR being a fixed, covalently-bound tetramer, these results suggest a more loosely-bound assembly of EGFR, comprising monomers and dimers that condense into clusters before EGF is added.

EGF causes clusters to increase their EGFR content. To determine the effect of EGF binding on EGFR cluster stoichiometry and spatiotemporal dynamics, we performed TIRF following addition of EGF to the cell culture. We kept live SW620-EGFR-GFP cells in serum-free media for 24 h prior to imaging to minimize binding of serum-based EGFR ligands and then washed immediately prior to EGF addition. We then added EGF conjugated 1:1 with fluorescent tetramethylrhodamine (EGF-TMR) at a final concentration of 100 ng/ml (15.6 nM), higher than the K_D for EGF to EGFR of 300 pM-2 nM (40), and visualized cells using TIRF to allow simultaneous observation of EGFR and EGF in our green- and red-colour channels, respectively. Excess EGF-TMR was retained in the sample during imaging enabling observation over incubation times of 3-60 min.

Colocalization of EGFR and EGF foci was determined using the numerical overlap integral between tracked green/red foci, establishing a metric for putative binding of EGF to EGFR clusters to within our spatial precision of 40 nm. After EGF incubation for a few minutes, binding between green/red foci was detected (Figure 3a, movie S2, Figure S8). We observed a mean of approximately 57 EGFR tracks per cell across all incubation times from 117 cells, and a total of 4,700 tracks across all cells (Table S1). We estimated 40±18% of EGFR clusters were bound to EGF over 3-60 min incubation, corresponding to 64% of all tracked EGFR clusters (Figure S9).

The EGFR stoichiometry for clusters not bound to EGF was similar to the value (~13 molecules) measured before adding EGF (Table S1, Figure 3b and Figure S9c). EGF-bound EGFR clusters had a higher mean stoichiometry of approximately 31 EGFR molecules compared to 11 EGFR molecules for clusters not bound to EGF, as shown in Table S1 and the stoichiometry

distributions in Figure 3b (Student's *t*-test P<0.0001), with non-parametric testing also indicating that the two distributions were statistically different (P<0.0001). Binning the stoichiometry as a function of incubation time (Figure S9c), the mean stoichiometry of EGFR clusters not bound to EGF remained roughly constant at 8-14 during incubation with EGF over 60 min, whereas that of EGF-bound EGFR clusters increased to 20-50 molecules per cluster.

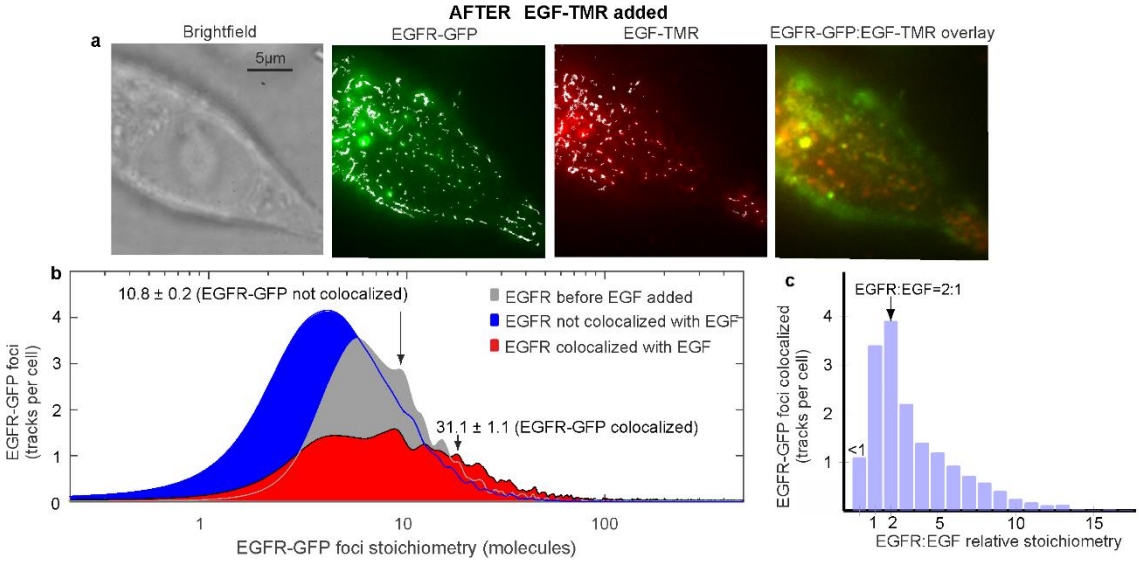


Figure 3. EGF increases EGFR cluster stoichiometry in SW620 cells. (a) *Brightfield and TIRF of SW620-EGFR-GFP cells after adding EGF-TMR (10 min time point). GFP (green), TMR (red) foci and overlay images are shown with yellow indicating colocalization (putative binding between EGFR clusters and EGF within our 40 nm spatial precision). Overlaid tracks shown (white).* ***(b)*** *Stoichiometry distributions of EGF-bound EGFR clusters (red) and EGFR not bound to EGF (blue) across all times. Mean and s.e.m. for each distribution indicated (arrows).* ***(c)*** *Distribution of relative EGFR:EGF stoichiometry for EGF-bound clusters. N=117 cells.*

EGF-bound clusters contain 4 EGFR molecules per EGF. To determine the relative stoichiometry between EGFR clusters and EGF when EGF was bound we measured red foci stoichiometry simultaneously to colocalized green foci. EGF stoichiometry was determined using the same photobleaching protocol to that of GFP-labeled EGFR. Fluorescence-intensity traces for EGF-TMR on the red channel exhibited step-wise photobleaching when multiple EGF-TMR molecules where present, and EGF foci stoichiometry was obtained by dividing the initial intensity in the traces by that of a single TMR molecule. As for GFP, the latter (approximately 2,400 counts on our detector) was obtained *in vivo* from the final brightness in the photobleach, averaging over multiple traces, and agreed with *in vitro* measurements (Figure S5c). Our analysis revealed a modal relative stoichiometry for EGFR:EGF of 1.9±0.8 (±half width half maximum, Figure 3c) with mean 4.2±0.1; EGFR clusters bound to EGF contain approximately 4 EGFR molecules for every EGF.

To interpret these observations, we developed a new multi-state time-dependent kinetics model that accounts for EGF-EGFR binding, receptor dimerization and receptor internalization and recycling (Figure S10a). The model predicts the fractional saturation on the surface, $Y^{surface}$, which is the surface ratio EGF:EGFR (excluding internalized molecules). The model shows that on adding EGF, initial concentrations of unligated EGFR monomers ([R]) and dimers ([RR]) decrease while concentrations of ligated monomers ([RL]) and dimers (singly ligated [RRL] and doubly ligated [RRL2]) increase over the first 5 min (Figure S10a). Endocytosis leads to accumulation of internalized ligated monomers ([RL^{inside}]) and dimers (singly ligated [RRL^{inside}] and doubly ligated dimers [$RRL2^{inside}$]) (dashed lines, Figure S10a) with EGFR recycling back to the plasma membrane contributing to equilibration of all concentrations after approximately

30 min (Figure S10a). $Y^{surface}$, is shown on the inset on Figure S10a. Its inverse at equilibrium predicts an EGFR:EGF ratio of ~1.5, lower than our observed mean ~4 value. However, if we assume that ligand can bind only to receptor monomers (and not to dimers), our model predicts $Y^{surface}$ of 0.24, which corresponds closely to the experimental mean EGFR:EGF ratio of ~4 (Figure S10c, d).

EGFR clustering increases on adding cetuximab or trastuzumab. It is known that EGF binds to monomeric EGFR resulting in EGFR dimerization prior to activation (5–8), however, it is less clear what role EGFR activation plays in EGFR clustering. To investigate the effect of EGFR pathway inhibition on EGFR clustering, we imaged the transfected SW620 cells in the presence of EGFR pathway inhibitors cetuximab or trastuzumab, two commonly used anti-cancer drugs, which separately target EGFR and HER2, respectively. These are, to our knowledge, the first single-molecule observations of the effect of EGFR- and HER2-targeting anti-cancer drugs on living human cancer cells. Cetuximab targets EGFR and is a monoclonal antibody anti-cancer drug commonly used against neck and colon cancers in advanced disease stages to inhibit cell division and growth (26). Binding of cetuximab to domain III of the soluble extracellular segment of EGFR is believed to result in partial blockage of the EGF-binding region, hindering the adoption of an extended conformation required for EGFR dimerization. Trastuzumab is a monoclonal antibody anti-cancer drug commonly used to treat breast cancer (27) that results in similar downstream effects of EGFR pathway inhibition of impairing cell division and growth. However, trastuzumab does not bind directly to EGFR but to domain IV of the extracellular segment of HER2 (41). Trastuzumab binding does not affect HER2 self-association (42) but influences the stability of HER2-mediated dimers with EGFR (43).

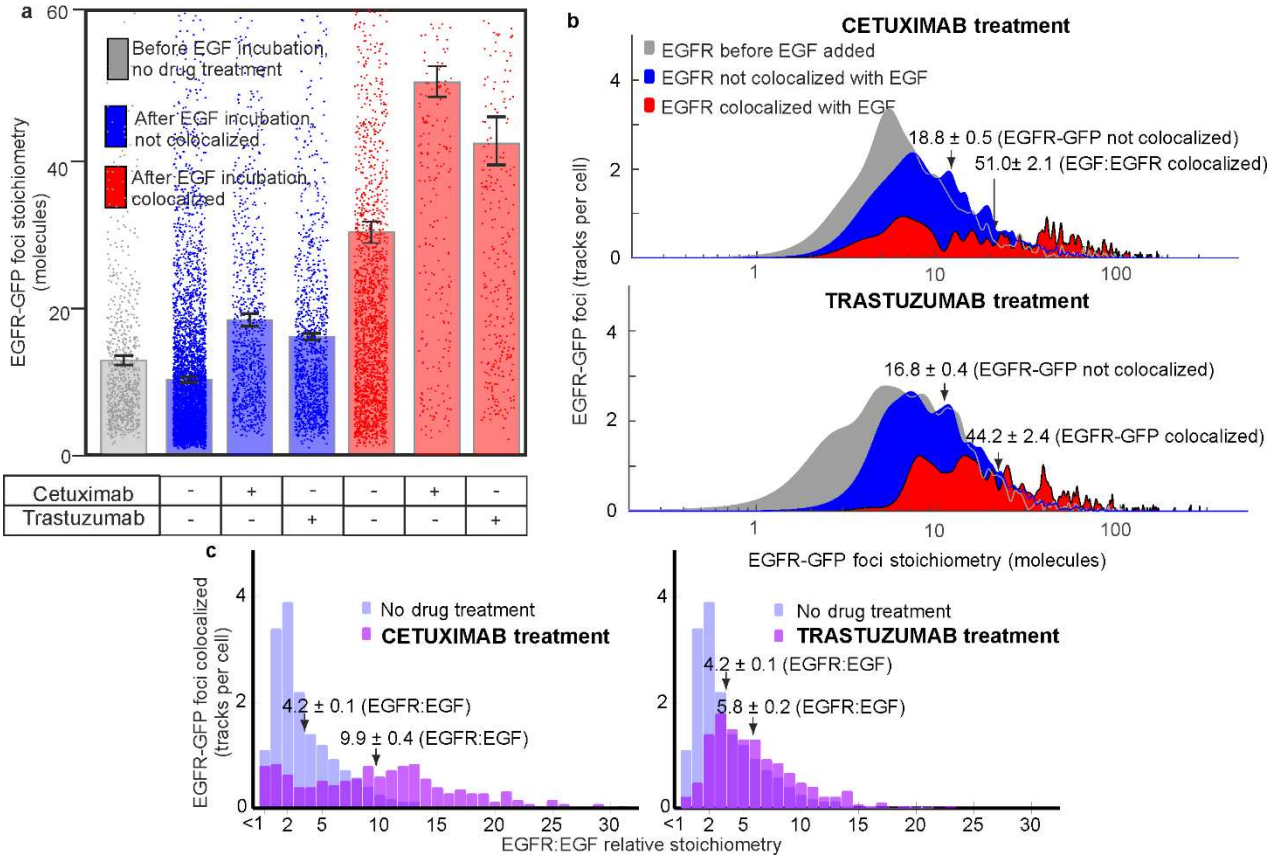


Figure 4. Cetuximab and trastuzumab increase EGFR cluster stoichiometry. ***(a)*** *Mean EGFR cluster stoichiometry before and after EGF incubation and its dependence on EGF binding, in the presence (+) or absence (-) of each drug treatment. Error bars are s.e.m, N =10-117 cells per dataset (see Table S1).* ***(b)****: Distributions of EGFR cluster stoichiometry for cells treated with*

cetuximab (top) or trastuzumab (bottom). Distributions shown are pre EGF addition (grey) and post EGF addition for EGF-bound EGFR clusters (red) and EGFR not bound to EGF (blue). Data collated across 60 min EGF incubation. Mean and s.e.m. values are indicated by arrows. ***(c)*** *Distributions of EGFR:EGF relative stoichiometries of EGF-bound EGFR clusters for drug-treated cells (purple) contrasted against no drug treatment (light blue). N=10-117 cells per dataset.*

Before adding EGF we found that treatment with cetuximab or trastuzumab at cytostatic concentrations caused statistically significant differences between the stoichiometry distributions for EGFR-GFP stoichiometry (Student's *t*-test, P<0.0001, Brunner-Munzel, P=0.01, P=0.08 respectively) but with no significant effect on the number of detected EGFR-GFP tracks per cell (Table S1). EGF incubation together with drug treatment resulted in increased EGFR cluster stoichiometries for both EGF-bound and EGFR-unbound clusters, for both drugs, compared to stoichiometries after EGF incubation with no drug treatment (Figure 4a,b; Table S1). The mean stoichiometry of EGF-bound EGFR clusters in drug-treated cells increased significantly to 51±2 and 44±2 for cetuximab and trastuzumab respectively, with maxima of several hundred molecules (Figs. 4a,b; Table S1). There were approximately 20% fewer EGF-bound EGFR tracks for cetuximab- or trastuzumab-treated cells compared to untreated cells (Figure S12). We also observed a shift to higher EGFR:EGF relative stoichiometry for cetuximab and trastuzumab treatments beyond the ~2:1 modal ratio observed for untreated cells (Figure 4c), consistent with competitive inhibition of EGF binding. Taken together, these results support the hypothesis that EGF binding increases the level of EGFR clustering.

EGF triggers larger EGFR heterocluster formation. Tracking of EGFR clusters indicated Brownian diffusion up to time intervals of approximately 100 ms (Figure S13). Using the initial gradient of the mean square displacement with respect to time interval for each track, we determined the apparent diffusion coefficient D and correlated this against EGFR cluster stoichiometry. Plotting D against stoichiometry for all tracked clusters shows a trend towards lower diffusion with higher stoichiometry (Figure 5a and Figure S13 and S14). One explanation for these observations can be made using the principles of the Stokes-Einstein relation, which states that $D=k_BT/\gamma$ where k_B is Boltzmann's constant, T is the absolute temperature and γ is the frictional drag of a tracked EGFR cluster in the membrane. The frictional drag is dependent on the local viscosity and the size and shape of the diffusing object. Larger clusters (i.e., those with a higher effective diameter) have a higher frictional drag in the membrane so a trend towards lower cluster diffusion with higher number of EGFR molecules per cluster is not unreasonable for an accretion model of cluster growth. In the absence of any drugs, D for EGF-bound clusters was lower than that for EGF-unbound clusters (red data to blue, Figure 5b), that would be consistent with an increase in effective cluster diameter as clusters accumulate more EGFR upon EGF binding triggering increased dimerization. However, for clusters that have grown much larger than the ~3-5 nm width of the 2D cell membrane, there is an expectation that the effective drag coefficient has a less-sensitive logarithmic scaling with effective diameter as opposed to being inversely proportional to the effective diameter of an object diffusing in a purely 3D environment (44), so there may be additional effects to consider (see Discussion).

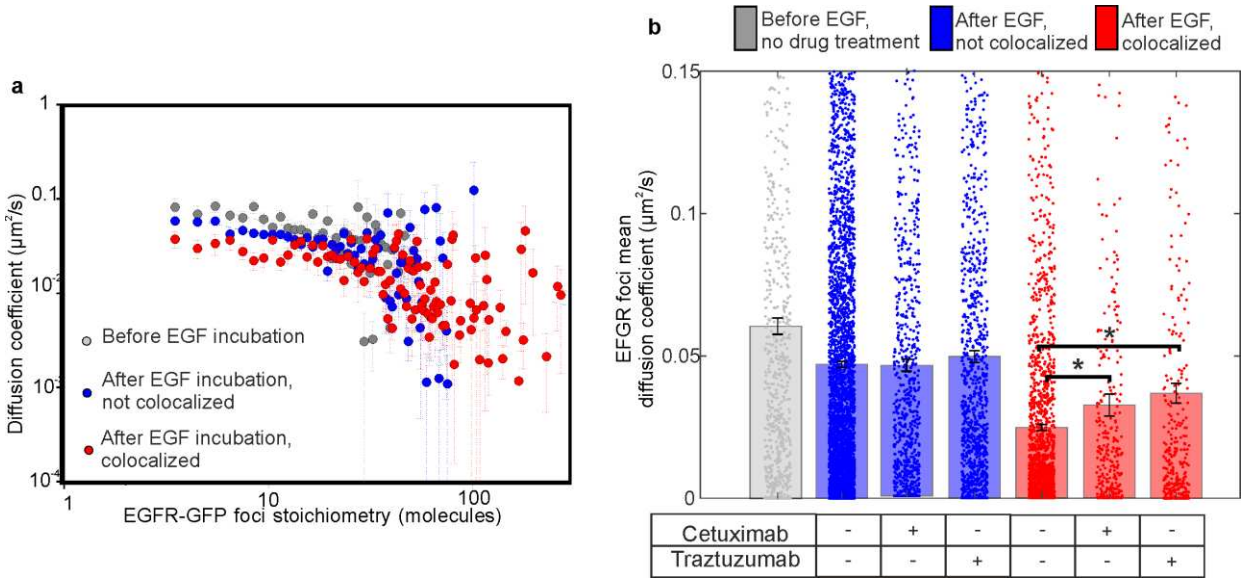


Figure 5. EGFR cluster diffusion depends on stoichiometry and EGF binding. ***(a)*** *log-log plot for apparent diffusion coefficient, D, as a function of EGFR cluster stoichiometry.* ***(b)*** *D upon drug treatment for the same datasets of Figure 4. Significant differences using Student's t-test ($P<0.05$) for +cetuximab and +trastuzumab ($P=0.01$ and <0.0001) are indicated with *, with corresponding Brunner-Munzel tests on the full distributions indicating P values of 0.0001 and 0.001, respectively; s.e.m. error bars.*

We found that addition of cetuximab or trastuzumab made relatively little difference to *D* for EGF-unbound EGFR clusters (blue data, Figure 5b), suggesting that these drug treatments are unlikely to have a significant effect on the plasma membrane viscosity in the vicinity of EGFR clusters, or on interactions of EGFR with the cytoskeleton that could influence cluster diffusion. However, we also found that both cetuximab and trastuzumab increased *D* for EGF-bound EGFR clusters (red data, Figure 5b) in the direction of higher values associated with EGFR clusters not bound to EGF in the untreated datasets (grey bar, Figure 5b). One hypothesis for these findings is that there are non-EGFR components present in clusters that influence *D*. HER2 is a candidate here, since trastuzumab binds not to EGFR but specifically to HER2; since the frictional drag of an EGFR cluster includes not only visible GFP-labeled EGFR but also any unlabeled components that contribute to frictional drag, one explanation is that trastuzumab reduces the EGFR cluster diameter by perturbing the association between EGFR and unlabeled HER2 if present in a cluster, following EGF binding to EGFR. In support of this explanation, HER2 is known to affect the stability of HER2-mediated dimers with EGFR (43) while not affecting the binding of HER2 with other HER2 molecules (42). An important conclusion to this hypothesis is that it is likely that, prior to drug treatment, there must clusters present that comprise both EGFR and HER2, i.e., heteroclusters.

A number of previous findings have inferred indirectly that EGFR may form heterocomplexes with other RTKs (14, 20, 22–25), and recent evidence shows that HER2 inhibitor lapatinib induces HER2/HER3 heterocomplex formation in breast cancer cells (45).

We have no available viable cell line derived from the SW620 cell line currently that has both HER2 and EGFR fluorescently labeled, however, we were able to construct a dual-label cell line using model CHO-K1 cells that have similar low endogenous EGFR expression levels. We constructed this cell line to contain GFP-labeled HER2 and EGFR labeled with HaloTag650 (HaloTag STELLA Fluor 650) ligand. Using similar TIRF and SMLM we found that HER2 and EGFR exhibit mobile and immobile foci, with transient colocalization and co-diffusion (Figure S15a) over a mean dwell time of 335±100 ms (Figure S15b, movies S3, S4). Although there are several biological differences between SW620 and CHO-K1 cells, taken together, our results add support to the hypothesis that EGF induces the formation of larger EGFR heteroclusters that

involve a HER2 component (Figure S15c). Heteroclusters may also include HER3 or HER4. We tested this indirectly by treating cells with the inhibitor pertuzumab, a monoclonal antibody anti-cancer drug similar to trastuzumab albeit with complementary function against HER2/HER3 heteroassociation (46). We found that pertuzumab treatment also resulted in perturbations to EGFR clustering (Figure S16), perhaps suggesting HER3 contribution to heteroclusters.

Discussion

Here, we investigated the role of EGFR clusters in cancer and their dependence on EGF binding. Two important improvements over earlier reports are: (i) our SW620 observations relate to a human carcinoma line, enabling insights to the EGF pathway in cancer directly; (ii) we have spatial information concerning EGFR and EGF localization simultaneously from labeled protein and ligand. In prior microscopy in which labeled EGF is not imaged simultaneously to labeled EGFR, inference is more limited.

We used single-molecule TIRF and SMLM on transfected SW620 cells which do not natively express EGFR. By using GFP on EGFR with TMR on EGF we have measured the stoichiometry and diffusion of single EGFR clusters, and how these depend on putative EGF binding is within our 40 nm spatial precision. We find that before EGF binds to EGFR, EGFR comprises clustered assemblies, the most prevalent of which contains 6 EGFR molecules, but with the cluster stoichiometry extending to several tens of molecules. We find that binding of EGF to EGFR results in higher cluster stoichiometry. The observation that EGFR may exist as pre-formed clusters prior to EGF activation has been suggested previously by several studies (11–15, 17) with our findings here adding to this growing consensus.

We developed a new time-dependent kinetics model using realistic parameters derived from previous studies. Unlike previous models, it predicts the time evolution of all concentrations and accounts for recycling and endocytosis. The model enables interpretation of imaging data revealing insights that could not be achieved with time-independent schemes based solely on affinities and equilibrium constants. It also factors in the temperature dependence of EGF-binding in living cells, showing the contrast between EGF:EGFR predictions at 37°C and 4°C ($Y^{surface}$~0.24 at 37°C versus $Y^{surface} \sim 0.96$ at 4°C). These differences arise from the fact that receptor-EGF binding and dimerization equilibrium constants can strongly depend on temperature (they may vary by as much as a factor of 10-100 between ~0°C and 37°C (47)), as well as from the fact that receptor internalization is highly temperature-dependent (48). It is worth noting that our model has validity for any receptor-ligand system for which reaction rates have been measured. It predicts a mean EGFR:EGF ratio of 4:1 within a cluster which agrees with our experimental measurements. The model predications are not explicitly dependent on the presence of heterodimers. Limited experimental data from heterodimeric components for EGF activation rates and internalisation processes preclude a full theoretical description within the current model framework, though it is not inconceivable that EGF-binding processes might be reaction-limited as opposed to diffusion-limited due to the relatively high rates of diffusion of the small EGF ligand comparable to EGFR clusters. In such a scenario, an EGFR molecule that is paired with aa non-EGFR ERBB superfamily partner (for example, HER2,3 or 4) might have comparable reaction kinetics to pure monomeric EGFR, though testing this is beyond the scope of our present study. To our knowledge, this is the first report of a truly time-dependent kinetics model applied to single-molecule precise live cell data, therefore its accurate prediction adds significant support to the model's key premise that there is preferential EGF binding to EGFR monomers with no binding to dimers.

Our model adds to existing evidence of ligand binding to EGFR monomers. Small angle X-ray scattering and isothermal titration calorimetry to EGFR's isolated extracellular domain (sEGFR) suggest EGF binds to sEGFR monomers, receptor dimerization involving association of two monomeric EGF-sEGFR (49). Multi-angle laser light scattering suggests sEGFR is monomeric in solution but dimeric after EGF ligation (50). Fluorescence anisotropy indicates 1:1 binding of EGF:sEGFR, analytical ultracentrifugation suggesting 2(EGF-sEGFR) complexes

(51). Structural evidence indicates activation is preceded by ligand binding to receptor monomers (52–54).

We also performed TIRF with SMLM to investigate anti-cancer drugs cetuximab (26) and trastuzumab (27), to our knowledge for the first time on living human cancer cells, although correlative fluorescence microscopy and liquid phase electron microscopy have been used previously to investigate a drug lapatinib that reversibly inhibits both EGFR and HER2 (55). We discovered that the diffusion of EGF-bound EGFR clusters increased upon treating cells with either drug. Since cetuximab and trastuzumab separately target EGFR and HER2 respectively (26, 41) a reasonable conclusion is that clusters likely contain a mixture of both proteins.

One implication for this observation is that these drugs reduce the effective frictional drag experience by clusters which could imply a compaction effect, i.e., that the EGFR packing density within EGF-bound clusters is higher for drug-treated compared to untreated cells. It is known that EGFR adopts a spatially more extended conformation for dimerization to occur (56) - when EGF binds to EGFR, the activated EGFR dimers become more compact than non-EGF-bound EGFR dimers in the 2D plane of the plasma membrane, but also become marginally taller perpendicular to this plane. Therefore, if a cluster contains a mixture of both EGF-bound and EGFR-unbound subunits then addition of a dimerization inhibitor might conceivably result in EGF-unbound EGFR subunits adopting the more compact conformation not associated with dimerization in the 2D plane of the plasma membrane, so increasing the overall packing density of EGFR in that cluster. However, equivalent details are not currently known for trastuzumab. An alternative explanation is that there are changes to the membrane or cytoskeletal microenvironment in the vicinity of EGFR clusters that are dependent not only on the presence of the drugs used here but also on whether EGF is bound to EGFR. One further consideration concerns putative hop diffusion that was reported as a model to explain the apparent increases in translational diffusion for E-cadherin oligomers in the plasma membrane (57). In this model, the plasma membrane is compartmentalized by the actin-based cytoskeleton into corral zones of a few hundred nm effective diameter that E-cadherin can hop between such that the hopping rate decreases dramatically with an increase of E-cadherin's oligomeric state. However, the spatial and temporal resolution limitations in our current work preclude us from probing this level of ultrastructural detail at this time.

Although we do not have a cell line in SW620 that co-expresses both fluorescently labeled EGFR and HER2, we were able to make a viable dual-label strain in model CHO-K1 cells, that indicated that EGFR and HER2 foci co-diffuse over periods of several hundred milliseconds prior to incubation with EGF. With the caveat that there are biological differences between SW620 and CHO-K1, if the molecular behavior of EGFR and HER2 molecular interactions per se are fundamentally identical irrespective of the cell line, these data suggest that EGFR clusters may contain a mixture of EGFR and HER2 both before and after EGF binding. It should be noted that other reports suggest that HER2 and HER3 may engage in heterocomplex formation (45) so we cannot exclude the possibility that HER3 may also be present in mixed clusters of the SW620 line. Indeed, we tested this possibility indirectly by treating cells with pertuzumab, a monoclonal antibody anti-cancer drug which targets the HER2/HER3 interface. This treatment also induced clustering, suggesting a role for HER3 in heteroclusters, however the full extent of HER3 and HER4 involvement in heteroclusters is beyond the scope of this present study.

Our findings show that EGFR is clustered before and after EGF binding, consistent with observations from AFM studies using EGF-coated tips which imaged human lung adenocarcinoma cells from the A549 cell line, known to have high EGFR expression (58). These data suggested half the EGFR clusters had diameters of 20-70 nm pre-activation, with 35-105 nm post activation, indicating cluster growth following EGF binding, to be compared with our findings. However, we find important differences with respect to some recent single-molecule studies. Although there were earlier suggestions of pre-formed oligomeric EGFR, including Needham *et al.* (17) and Zanetti-Domingues *et al.* (59), they and Huang *et al.* (10), observed

monomeric EGFR, in particular Huang *et al.* assign a proportion of 94%. We cannot directly exclude the possibility in our experiments that monomeric EGFR is at high density in for which the mean separation is less than the optical resolution limit. However, the absence of not a single detected monomer from several thousand tracks, despite having single GFP detection sensitivity, makes this unlikely. An alternative explanation is that we estimate the EGFR copy number to be ~200,000 molecules per cell, similar to endogenously expressing cancer cell lines (60) but more than double that estimated from Needham *et al.* and Huang *et al.*, which may conceivably result in shifting the equilibrium position for EGFR clustering towards higher stoichiometry. In support of this, the peak value of 6 EGFR cluster molecules we measure before EGF binds is higher than that observed by Needham *et al.* who observed 4. Such an increased on-rate could conceivably contribute to a depleted monomeric population, which has implications for carcinomas in which the expression level of EGFR is known to be high. Absence of monomeric EGFR before addition of EGF may also suggest some spontaneous activation independent of ligand binding.

Reports on possible heterocomplex formation between EGFR and other ERBB proteins do not detail whether these associations are before or after EGF binding. Our observations show that transient associations between EGFR and HER2 may last a few hundred milliseconds, but that cluster size and number increase following EGF binding. Our findings suggest a role for trastuzumab in modulating regulatory balance through the availability of endogenous HER2 to associate with EGFR. Even when scarce, the presence of HER2 is known to selectively discourage internalization and degradation of activated EGFR, and promote recycling to the plasma membrane both via chaperone proteins and EGF dissociation (61). However, although HER2 is known to act as co-receptor, it has no known direct ligand and its physiological role in interacting with EGFR is still unclear. One possibility is that the diffusion of heterocomplexes may enable a spread of activation across cell surfaces. Also, the resistance of HER2-bearing complexes to downregulation might sustain signaling once established, i.e. a “latch” response. Future work involving development of a viable SW620 cell line that co-expresses labeled EGFR and HER2 may help these questions to be addressed, in particular to determine what ERBB component EGF specifically targets in clusters that contain heterodimers.

Future work will also be valuable to unravel how EGFR/HER2 heterocluster formation affects and is affected by the downstream signaling proteins, which themselves may cluster and alter their diffusion as has been observed in Ras signaling which interacts with EGFR.(62) Similar bidirectional affects occur with the cytoskeleton and through endocytosis. Constraining EGFR clustering and diffusion modulates phosphorylation (63), similarly inhibiting endocytosis increases EGFR autophosphorylation (64). Unravelling the complex interplay between receptor clustering and diffusion with downstream signaling proteins, cytoskeletal interactions and endocytosis will remain a significant challenge going forward.

Our findings that heterocomplex cluster size increase post EGF binding suggest new strategies for anti-cancer drug design. For example, new drugs to disrupt interfaces between HER2 and EGFR directly. Strategies that disrupt EGFR clusters before EGF binding may also inspire new drug designs. Similarly, single-molecule quantification would be valuable to probe different carcinomas, for example those of the lung in which EGFR mutations are implicated in cancer (65) or in the design of cell-surface logic gates for targeted therapies (66). Finally, in enabling quantification of the actions of different drugs, there may be value in identifying chemotherapy “sweet-spots” in carcinomas known to be treatable using combined drugs, such as in gastric cancer (67).

Materials and Methods

Full details for methods used for cell line preparation, gene expression quantification, microscopy and modelling given in Supplementary.

Software access. All bespoke code in MATLAB is available EGFRanalyser https://sourceforge.net/projects/york-biophysics/.

Data availability. We do not upload additional data analysis files since analyzed data are included in full in the main text and supplementary files. All raw imaging data are available from the authors.

Acknowledgements: Thanks Philippe Bastiaens, Max Planck Institute of Molecular Physiology, Dortmund, Germany for donation of perbB1-EGFP-N1, Ivan R. Nabi, University of British Columbia, Canada for donation of EGFR-YFP plasmid, Taka-aki Tsunoyama, OIST, Japan for training and discussions, and Hannah Walker and Norman Maitland for technical advice concerning cancer cell maintenance and western blotting and for resources at the Cancer Research Unit, University of York.

Funding: Work was supported by the EPSRC (EP/G061009/1), Royal Society (RG0803569, UF110111), BBSRC (BB/F021224/1, BB/N006453/1), MRC (MR/K01580X/1), JSPS (International Research Fellowship PE18730), and CRUK (C38302/A12278).

Author contributions: CF, DO, JW, SS, created and biologically characterized the cell line. OH, ILG built the microscope. CF, ILG, OH, AL, PZ collected the microscopy data. AW, CF wrote analysis software. AW, CF, AH, PZ, TCL analyzed the data. ILG performed modeling. WB and MCL designed the study. All authors wrote the manuscript.

Competing interests: We declare no competing interests.

References

1. R. Roskoski, The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol. Res.* **79**, 34–74 (2014).
2. R. N. Jorissen, *et al.*, Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp. Cell Res.* **284**, 31–53 (2003).
3. I. Lax, *et al.*, Functional analysis of the ligand binding site of EGF-receptor utilizing chimeric chicken/human receptor molecules. *EMBO J.* **8**, 421–427 (1989).
4. M. R. Schneider, E. Wolf, The epidermal growth factor receptor ligands at a glance. *J. Cell. Physiol.* **218**, 460–466 (2009).
5. K. M. Ferguson, *et al.*, EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol. Cell* **11**, 507–517 (2003).
6. J. L. Macdonald-Obermann, L. J. Pike, The Intracellular Juxtamembrane Domain of the Epidermal Growth Factor (EGF) Receptor Is Responsible for the Allosteric Regulation of EGF Binding. *J. Biol. Chem.* **284**, 13570–13576 (2009).
7. P. Liu, *et al.*, A single ligand is sufficient to activate EGFR dimers. *Proc. Natl. Acad. Sci.* **109**, 10861–10866 (2012).
8. J. L. Macdonald, L. J. Pike, Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proc. Natl. Acad. Sci.* **105**, 112–117 (2008).
9. Y. Sako, S. Minoghchi, T. Yanagida, Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat. Cell Biol.* **2**, 168–72 (2000).
10. Y. Huang, *et al.*, Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *Elife* **5**, e14107 (2016).
11. J. Ichinose, M. Murata, T. Yanagida, Y. Sako, EGF signalling amplification induced by dynamic clustering of EGFR. *Biochem. Biophys. Res. Commun.* **324**, 1143–1149 (2004).
12. M. Martin-Fernandez, D. T. Clarke, M. J. Tobin, S. V Jones, G. R. Jones, Preformed oligomeric epidermal growth factor receptors undergo an ectodomain structure change during signaling. *Biophys. J.* **82**, 2415–2427 (2002).
13. A. H. A. Clayton, *et al.*, Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis.

J. Biol. Chem. **280**, 30392–30399 (2005).
14. R.-H. Tao, I. N. Maruyama, All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells. *J. Cell Sci.* **121**, 3207–3217 (2008).
15. P. Nagy, J. Claus, T. M. Jovin, D. J. Arndt-Jovin, Distribution of resting and ligand-bound ErbB1 and ErbB2 receptor tyrosine kinases in living cells using number and brightness analysis. *Proc. Natl. Acad. Sci.* **107**, 16524–16529 (2010).
16. N. Kozer, *et al.*, Exploring higher-order EGFR oligomerisation and phosphorylation - A combined experimental and theoretical approach. *Mol. Biosyst.* **9**, 1849–1863 (2013).
17. S. R. Needham, *et al.*, EGFR oligomerization organizes kinase-active dimers into competent signalling platforms. *Nat. Commun.* **7**, 13307 (2016).
18. D. B. Peckys, U. Korf, N. De Jonge, Local variations of HER2 dimerization in breast cancer cells discovered by correlative fluorescence and liquid electron microscopy. *Sci. Adv.* **1** e1500165 (2015).
19. S. J. Tobin, *et al.*, Single molecule localization microscopy coupled with touch preparation for the quantification of trastuzumab-bound HER2. *Sci. Rep.* **8**, 1–13 (2018).
20. M. R. Siddiqui, *et al.*, Targeting Epidermal Growth Factor Receptor (EGFR) and Human Epidermal Growth Factor Receptor 2 (HER2) Expressing Bladder Cancer Using Combination Photoimmunotherapy (PIT). *Sci. Rep.* **9**, 1–14 (2019).
21. A. SQ, *et al.*, Direct and immune mediated antibody targeting of ERBB receptors in a colorectal cancer cell-line panel. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21046–21051 (2012).
22. F. Weinberg, D. B. Peckys, N. de Jonge, Egfr expression in her2-driven breast cancer cells. *Int. J. Mol. Sci.* **21**, 1–19 (2020).
23. I. N. Dahmke, *et al.*, Correlative Fluorescence- and Electron Microscopy of Whole Breast Cancer Cells Reveals Different Distribution of ErbB2 Dependent on Underlying Actin. *Front. Cell Dev. Biol.* **8** 521 (2020).
24. R. H. Tao, I. N. Maruyama, All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells. *J. Cell Sci.* **121**, 3207–3217 (2008).
25. M. L. I. E. Harwardt, *et al.*, Single-molecule super-resolution microscopy reveals heteromeric complexes of MET and EGFR upon ligand activation. *Int. J. Mol. Sci.* **21** 2803 (2020).
26. P. Kirkpatrick, J. Graham, M. Muhsin, Fresh from the pipeline: Cetuximab. *Nat. Rev. Drug Discov.* **3**, 549–550 (2004).
27. K. P. Garnock-Jones, G. M. Keating, L. J. Scott, Trastuzumab. *Drugs* **70**, 215–239 (2010).
28. A. Leibovitz, *et al.*, Classification of Human Colorectal Adenocarcinoma Cell Lines. *Cancer Res.* **36** (1976).
29. J. L. Wilding, S. McGowan, Y. Liu, W. F. Bodmer, Replication error deficient and proficient colorectal cancer gene expression differences caused by 3'UTR polyT sequence deletions. *Proc. Natl. Acad. Sci.* **107**, 21058–21063 (2010).
30. K. S. Frederiksen, N. Abrahamsen, R. J. Cristiano, L. Damstrup, H. S. Poulsen, Gene delivery by an epidermal growth factor/DNA polyplex to small cell lung cancer cell lines expressing low levels of epidermal growth factor receptor. *Cancer Gene Ther.* **7**, 262–268 (2000).
31. L. Dahan, A. Sadok, J. L. Formento, J. F. Seitz, H. Kovacic, Modulation of cellular redox state underlies antagonism between oxaliplatin and cetuximab in human colorectal cancer cell lines. *Br. J. Pharmacol.* **158**, 610–620 (2009).
32. Y.-H. Shih, *et al.*, 111 In-cetuximab as a diagnostic agent by accessible epidermal growth factor (EGF) receptor targeting in human metastatic colorectal carcinoma. *Oncotarget* **6**, 16601–16610 (2015).
33. J.-L. Yang, X.-J. Qu, P. J. Russell, D. Goldstein, Regulation of epidermal growth factor receptor in human colon cancer cell lines by interferon a. *Gut* **53** 123–9 (2004)
34. R. J. Coffey, *et al.*, Transforming Growth Factor α and β Expression in Human Colon Cancer Lines: Implications for an Autocrine Model. *Cancer Res.* **47** 4590–4 (1987).

35. M. C. Leake, *et al.*, Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* **443**, 355–358 (2006).
36. V. Palmieri, *et al.*, Mechanical and structural comparison between primary tumor and lymph node metastasis cells in colorectal cancer. *Soft Matter* **11**, 5719–5726 (2015).
37. A. J. M. Wollman, M. C. Leake, Millisecond single-molecule localization microscopy combined with convolution analysis and automated image segmentation to determine protein concentrations in complexly structured, functional cells, one cell at a time. *Faraday Discuss.* **184**, 401–24 (2015).
38. I. Llorente-Garcia, *et al.*, Single-molecule in vivo imaging of bacterial respiratory complexes indicates delocalized oxidative phosphorylation. *Biochim. Biophys. Acta* **1837**, 811–24 (2014).
39. S. R. Needham, *et al.*, Measuring EGFR separations on cells with ~10 nm resolution via fluorophore localization imaging with photobleaching. *PLoS One* **8**, e62331 (2013).
40. M. A. Lemmon, Ligand-induced ErbB receptor dimerization. *Exp. Cell Res.* **315**, 638–648 (2009).
41. H.-S. Cho, *et al.*, Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**, 756–760 (2003).
42. H. Maadi, B. Nami, J. Tong, G. Li, Z. Wang, The effects of trastuzumab on HER2-mediated cell signaling in CHO cells expressing human HER2. *BMC Cancer* **18**, 238 (2018).
43. T. S. Wehrman, *et al.*, A system for quantifying dynamic protein interactions defines a role for Herceptin in modulating ErbB2 interactions. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19063–19068 (2006).
44. P. G. Saffman, M. Delbrück, Brownian motion in biological membranes. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 3111–3 (1975).
45. J. Claus, *et al.*, Inhibitor-induced HER2-HER3 heterodimerisation promotes proliferation through a novel dimer interface. *Elife* **7**, e32271 (2018).
46. J. Rockberg, J. M. Schwenk, M. Uhlén, Discovery of epitopes for targeting the human epidermal growth factor receptor 2 (HER2) with antibodies. *Mol. Oncol.* **3**, 238–247 (2009).
47. E. C. Hulme, M. A. Trevethick, Ligand binding assays at equilibrium: Validation and interpretation. *Br. J. Pharmacol.* **161**, 1219–1237 (2010).
48. A. Sorkin, J. E. Duex, Quantitative analysis of endocytosis and turnover of epidermal growth factor (EGF) and EGF receptor. *Curr. Protoc. cell Biol.* **Chapter 15**, Unit 15.14 (2010).
49. M. A. Lemmon, *et al.*, Two EGF molecules contribute additively to stabilization of the EGFR dimer. **16**, 281–294 (1997).
50. M. Odaka, D. Kohda, I. Lax, J. Schlessinger, F. Inagaki, Ligand-binding enhances the affinity of dimerization of the extracellular domain of the epidermal growth factor receptor. *J. Biochem.* **122**, 116–21 (1997).
51. T. Domagala, *et al.*, Stoichiometry, kinetic and binding analysis of the interaction between epidermal growth factor (EGF) and the extracellular domain of the EGF receptor. *Growth Factors* **18**, 11–29 (2000).
52. H. Ogiso, *et al.*, Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110**, 775–787 (2002).
53. T. P. J. Garrett, *et al.*, Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* **110**, 763–773 (2002).
54. H.-S. Cho, D. J. Leahy, Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* **297**, 1330–3 (2002).
55. F. Weinberg, D. B. Peckys, N. de Jonge, Egfr expression in her2-driven breast cancer cells. *Int. J. Mol. Sci.* **21**, 1–19 (2020).
56. A. Arkhipov, *et al.*, Architecture and Membrane Interactions of the EGF Receptor. *Cell*

152, 557–569 (2013).
57. R. Iino, I. Koyama, A. Kusumi, Single molecule imaging of green fluorescent proteins in living cells: E-cadherin forms oligomers on the free cell surface. *Biophys. J.* **80**, 2667 (2001).
58. W. Zhao, *et al.*, Mapping the resting and stimulated EGFR in cell membranes with topography and recognition imaging. *Anal. Methods* **6**, 7689–7694 (2014).
59. L. C. Zanetti-Domingues, *et al.*, The architecture of EGFR's basal complexes reveals autoinhibition mechanisms in dimers and oligomers. *Nat. Commun.* **9**, 4325 (2018).
60. F. Zhang, *et al.*, Quantification of Epidermal Growth Factor Receptor Expression Level and Binding Kinetics on Cell Surfaces by Surface Plasmon Resonance Imaging. *Anal. Chem.* **87**, 9960–9965 (2015).
61. Z. Wang, L. Zhang, T. K. Yeung, X. Chen, Endocytosis Deficiency of Epidermal Growth Factor (EGF) Receptor–ErbB2 Heterodimers in Response to EGF Stimulation. *Mol. Biol. Cell* **10**, 1621–1636 (1999).
62. R. Nussinov, C. J. Tsai, H. Jang, Oncogenic KRas mobility in the membrane and signaling response. *Semin. Cancer Biol.* **54**, 109–113 (2019).
63. D. Stabley, S. Retterer, S. Marshall, K. Salaita, Manipulating the lateral diffusion of surface-anchored EGF demonstrates that receptor clustering modulates phosphorylation levels. *Integr. Biol.* **5**, 659–668 (2013).
64. L. P. Sousa, *et al.*, Suppression of EGFR endocytosis by dynamin depletion reveals that EGFR signaling occurs primarily at the plasma membrane **109** 4419–24 (2012).).
65. J. G. Paez, *et al.*, EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
66. M. J. Lajoie, *et al.*, Designed protein logic to target cells with precise combinations of surface antigens. *Science* **369** 1637–1643 (2020).
67. K. Aoyagi, *et al.*, Molecular targeting to treat gastric cancer. *World J. Gastroenterol.* **20**, 13741–13755 (2014).
68. A. J. Wollman, *et al.*, Transcription factor clusters regulate genes in eukaryotic cells. *Elife* **6** (2017).
69. S. Sigismund, *et al.*, Clathrin-independent endocytosis of ubiquitinated cargos. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2760–2765 (2005).
70. P. J. Enriori, *et al.*, Breast cyst fluids increase the proliferation of breast cell lines in correlation with their hormone and growth factor concentration. *Clin. Endocrinol. (Oxf).* **64**, 20–28 (2006).
71. X. Michalet, Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E* **82**, 041914 (2010).
72. K. G. N. Suzuki, *et al.*, Transient GPI-anchored protein homodimers are units for raft organization and function. *Nat. Chem. Biol.* **8**, 774–83 (2012).
73. A. Kusumi, *et al.*, Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: High-speed single-molecule tracking of membrane molecules. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 351–378 (2005).
74. J. Barretina, *et al.*, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
75. M. R. Holbrook, J. B. O'Donnell, L. L. Slakey, D. J. Gross, Epidermal growth factor receptor internalization rate is regulated by negative charges near the SH2 binding site Tyr992. *Biochemistry* **38**, 9348–56 (1999).
76. S. Felder, J. Lavin, A. Ullrich, J. Schlessinger, Kinetics of Binding, Endocytosis, and Recycling of EGF Receptor Mutants. *J. Cell Biol.* **117**, 203–212 (1992).
77. R. S. Kasai, *et al.*, Full characterization of GPCR monomer-dimer dynamic equilibrium by single molecule imaging. *J. Cell Biol.* **192**, 463–480 (2011).
78. E. C. Hulme, M. A. Trevethick, Ligand binding assays at equilibrium: Validation and

interpretation. *Br. J. Pharmacol.* **161**, 1219–1237 (2010).
79. M. C. Leake, D. Wilson, B. Bullard, R. M. Simmons, The elasticity of single kettin molecules using a two-bead laser-tweezers assay. *FEBS Lett.* **535**, 55–60 (2003).
80. M. C. Leake, D. Wilson, M. Gautel, R. M. Simmons, The elasticity of single titin molecules using a two-bead optical tweezers assay. *Biophys. J.* **87**, 1112–35 (2004).

Supplementary Information

Supplementary Methods:

Cell lines. From ~100 screened colon carcinoma lines SW620, COLO320HSR and COLO741 all exhibited low endogenous EGFR expression from microarray data (29) (Figure S2) COLO741 was found to be a melanoma and COLO320HSR exhibited transfection instability so SW620 was selected. EGFR protein was undetectable in SW620 by western blotting (Figure S1a), confocal fluorescence microscopy (Figure S1b) and mRNA microarray (Figure S2a). Lysate protein levels were estimated by radio immunoprecipitation and BCA (Thermo Scientific™ Pierce™). Western blotting was performed using anti-EGFR mouse monoclonal (1:1000, clone 1F4, Cell Signaling Technology®) and anti-β-tubulin mouse monoclonal (1:1000, Sigma-Aldrich®) antibodies prepared in TBS-T, 5% milk, incubated overnight at 4ºC, then incubated with secondary antibody of polyclonal rabbit anti-mouse conjugated to horseradish peroxidase (Dako) at 1:10,000 and 1:100,000 for EGFR and β-tubulin, respectively, prior to chemiluminescence exposure (Amersham Biosciences). Cell lines with intermediate EGFR expression were used as positive controls.

Plasmid perbB1-EGFP-N1 (donated, Philippe Bastiaens) was used for transformations, comprising human *EGFR* insertion into enhanced GFP backbone pEGFPN1 plus kanamycin resistance. This GFP had no effect on protein clustering compared to the A206K mutant variant in a previous single-molecule study (68). 1 day pre-transfection, 200,000 SW620 cells in 1 ml growth media were seeded into a 12-well plate, adding 2 μg of pEGFR-EGFP (Invitrogen), 200 μl Invitrogen Opti-MEM® I Reduced Serum Medium, 1 μl of Plus™ Reagent and 6 μl of Lipofectamine® LTX to each. DNA-lipid complexes were added dropwise to cells, incubated 5 h (5% CO_2, 37 ºC), then media exchanged to normal the following day. Cells were reseeded onto 15 cm plates in Gibco® Dulbecco's modified eagle media (DMEM) supplemented with 4.5 g/l glucose, pyruvate, L-glutamine and phenol red plus 2 μg/ml Gibco™ Geneticin® (G418 sulfate). Colonies were isolated using a silicon-cloning cylinder (Corning®), harvested by trypsinization and transferred in a 12-well plate. Flow cytometry-based fluorescence assisted cell sorting (FACS) was used to sort cells into three fractions based on fluorescence intensity using the lowest intensity fraction (i.e., lowest EGFR-GFP expression) for subsequent experiments to minimize effects of expression variability across the cell population. Transgene expression was confirmed by mRNA microarray (Figure S2b), imaging live and immunofluorescently stained fixed cells with confocal microscopy, and western blotting. Confocal fluorescence microscopy confirmed EGFR plasma membrane localization (Figure S1b) and immunofluorescence on fixed cells using AlexaFluor633-labelled anti-EGFR and anti-GFP antibodies further demonstrated colocalization of GFP and EGFR (Figure S3).

EGFR-GFP kinase activity in SW620 cells was confirmed by observing increased phosphorylation of EGFR downstream targets, ERK1/2, in response to EGF (Figure S2b). We used an EGF concentration (either unlabeled or as EGF-RMR in dual-colour TIRF microscopy) equivalent to 100 ng/ml. This level resulted in clear phosphorylation activity on western blots. It is consistent with high physiological levels found in prostate and breast tissue – in particular 50-500 ng/mL found in high EGF bodily fluids including prostate fluid (69) or 30-300 ng/mL found in breast cysts (70).

RT-qPCR indicated low endogenous expression of HER2, HER3 and HER4, and that EGFR-GFP transfection did not cause significant changes in the expression of HER2 and HER3 (Figure S2c).

For CHO-K1, cDNA encoding EGFR tagged with EGFR-Halo was generated by replacing cDNA encoding YFP protein in human EGFR-YFP plasmid (donated, Ivan Nabi), with that of

Halo 7-tag protein (Promega), with insertion of a 45-base linker (15 aa, sequence 3SGGG) between EGFR and Halo. cDNA encoding human Erb2 (National Institute of Technology and Evaluation Biological Resource Centre 2-49-10, Nishihara,Shibuya-ku,Tokyo 151-0066 Japan) was fused at its C-terminus with mGFP (EGFP with A206K mutation), placed in the pOSTet153T vector, inserting an 18-base linker (6 aa, with sequence of three SG repeats) between Erb2 and mGFP. All constructs were confirmed by DNA sequencing.

Fab. IgG antibodies to EGF and anti-EGF rabbit anti-mouse polyclonal IgG (Molecular Probes) were digested by papain, confirmed by migration of 28-30 kDa and 25 kDa proteins corresponding to reduced Fc and Fab respectively. Fab was purified using protein A immobilized within a spin column, evaluated by 280 nm absorbance (Thermo Scientific NanoDrop).

Confocal. Zeiss inverted Axio Observer Z1 microscope with LSM 510 META scanning module and Plan-Aprochromat 63x 1.40NA oil immersion DIC M27 objective lens was used, enabling simultaneous imaging of green/red channels via 488 nm/565 nm wavelengths. SW620:EGFR-GFP cells grown in Corning 75 cm^2 treated plastic cell culture flasks in a humidified incubator (37 ºC, 5% CO_2) once 70-100% confluent were sub-cultured by trypsinization. 2-7 days prior to imaging, ~200,000 cells were seeded onto a Ibidi µ-dish 35 mm, high glass bottom using their normal culture media, DMEM, containing phenol red, then changed to DMEM with addition of 4.5 g/l glucose, L-glutamine, HEPES, without phenol red, and supplemented with 10% FBS, 100 units/ml penicillin and 100 µg/ml streptomycin, or directly into DMEM without phenol red as appropriate. Prior to imaging media was changed to Molecular Probes® Live Cell Imaging Solution supplemented with 1.5 mg/ml G418 sulfate.

For immunofluorescence we harvested SW620-EGFR-GFP cells 48 h prior to fixation at ~50,000 density per well seeded into Ibidi µ-Slide VI0.4, cultured in DMEM without phenol red, supplemented with 4.5 g/l glucose, L-glutamine, HEPES, 10% FBS and 100 units/ml of penicillin and 100 µg/ml streptomycin, 1.5 mg/ml G418. Cells were fixed with 4% formaldehyde at room temperature for 10min and washed. Non-specific antibody adsorption was blocked with 10% FBS in PBS for 10-20min. Primary antibodies were EGFR (D38B1) XP rabbit monoclonal 4267P (Cell Signaling Technology, 1:50 dilution) and anti-GFP chicken IgY (H+L) (Cell Signaling Technology, 1:400 dilution) in PBS with 10% FBS and 0.1% saponin overnight at 4 ºC. Each well was washed with 10% FBS and incubated with secondary antibodies, DyLight 633 goat anti-rabbit immunoglobulin G (IgG) highly cross adsorbed (PN35563, Thermo Scientific), 1:200, and Alexa Fluor 633 goat anti-chicken IgG (H+L) 2 mg/ml (Invitrogen) in PBS with 10% FBS and 0.1% saponin. Channels were washed with PBS and Sigma Aldrich Mowiol 4-88 added to solidify overnight. GFP, DyLight 633 or Alexa Fluor 633 and 4',6-diamidino-2-phenylindole (DAPI) were individually illuminated and scanned (indicating no mycoplasma). GFP was excited as for live cell imaging, while DyLight 633 and Alexa Fluor 633 were excited by a 633 nm HeNe laser.

TIRF. For SW620:EGFR-GFP, a dual-colour single-molecule microscope was modified from previous designs (38) equipped with a nanostage (Mad City Labs) and a 37 ºC humidified incubator supplemented with 5% CO_2 (INUB-LPS, Tokai Hit) (Figure S4) We used Elforlight B4-40 473 nm 40 mW and Oxxius SLIM 561 nm 200 mW lasers attenuated into a common path prior to polarization circularization (achromatic λ/4 plate) before entering a Nikon Eclipse-Ti inverted microscope body. An achromatic lens mounted onto a translation stage controlled the angle of incidence into the objective lens to generate TIRF via a Semrock 488/561nm BrightLine® dual-edge laser-flat dichroic beam splitter into a Nikon TIRF 100x NA1.49 oil immersion objective lens enabling simultaneous GFP/TMR detection across a 20µm full width at half maximum field, intensity 1kW/cm^2, 100nm penetration depth. Fluorescence was sampled 30ms per frame imaging onto two 512x512 pixel array EMCCD cameras (Andor, iXon+ DU-897 and iXon DU-887 for green/red, piezoelectrically cooled to -70ºC), 50nm/pixel magnification, via Semrock 561nm StopLine® single notch and Chroma 473 nm notch filters. Typically, scans were 200 frames.

For *in vitro* TIRF we used surface-immobilized GFP or EGF-TMR via anti-GFP or anti-EGF antibodies (Molecular Probes) or Fab followed by BSA passivation prior to washing (35). Slides were constructed from Ibidi sticky-Slides VI0.4 and 25 mm×75 mm No. 1.5 D263M Schott plasma-cleaned glass coverslip and IgG/Fab applied to a single channel and incubated at room temperature for 5min, washed x3 PBS, blocked with 1mg/ml of BSA for 60 min. The channel was again washed x3 then incubated with GFP for 7.5 min or EGF-TMR for 4 min. The channel was washed x5 before adding 1:10000, 200 nm diameter, 4% w/v, Invitrogen Molecular Probes carboxyl latex beads for focusing.

For live cell TIRF, cells were seeded/grown in media onto glass-bottomed Petri dishes or Corning culture flasks at 37 ºC, 5% CO_2. SW620:EGFR-GFP, or SW620 as negative control, imaged on either i) plasma cleaned glass coverslips (25 mm×75 mm No. 1.5 D263M Schott) covered by a sterile Ibidi sticky-Slide VI0.4, or ii) Ibidi µ-dish 35 mm, high glass bottom as for confocal. 48 h prior to imaging, cells were seeded onto the imaging chamber at ~200,000/cm^2 density. For slides, 50 µl (or 800 µl for dishes) DMEM without phenol red supplemented with 10% FBS, 100 units/ml penicillin and 100µg/ml streptomycin was added. 24h prior to imaging media was changed to DMEM without phenol red supplemented with 100 units/ml penicillin, 100µg/ml streptomycin and 1.5mg/ml G418 sulfate plus inhibitors (2ng/ml cetuximab (BioVision), 10ng/ml trastuzumab (BioVision) or 20ng/ml pertuzumab (Selleck Chemicals)) where necessary, without FBS (starving cells of residual serum EGF) for 24 h. We checked SW620 for expression of the most common ligands, using publicly available RNA-Seq data and our microarray data: EGF zero; TGFA low level; HBEGF low level expression; AREG zero; BTC zero; EREG zero; EPGN no data available (Supplementary). Although we cannot rule out the presence of very low levels of TGFA, cells were washed prior to imaging and no change was observed in EGFR clustering over 60 min unless EGF was added (Figure 3) suggesting no or negligible autocrine EGFR stimulation. Immediately before imaging, media was exchanged to Molecular Probes® Live Cell Imaging Solution supplemented with G418 sulfate and inhibitors where appropriate. Fluorescence sequences at 5min intervals up to 60 min were acquired after adding 100 ng/ml (15.6 nM) EGF-TMR (Molecular Probes).This EGF concentration resulted in clear phosphorylation activity on western blots and is consistent with high physiological levels found in prostate and breast tissue (Figure S2b).

CHO-K1 cells were illuminated using a different TIRF microscope with similar capability. Objective lens-based excitation was used with an evanescent field of 100 nm, and 37˚C stage temperature control, around an IX-83, Olympus inverted microscope with Olympus 100× NA1.49 oil immersion objective lens, laser powers 1.2 mW and 5 mW for 488 nm and 642 nm lasers. Dual-colour images were separated by dichroic mirrors (ZT405/488/561/640rpc-UF3, ZT561rpc-UF3 and ZT640rpc-UF3; Chroma), projected into green/red detection channels with emission filters of 500–550 nm for HER2-mGFP (ET525/50m; Chroma) and 662.5–737.5 nm (ET525/50m; Chroma) for EGFR labelled with HaloTag STELLA Fluor 650 ligand (a red fluorescent dye), then onto a two-stage microchannel plate intensifier (C9016-02MERLP24; Hamamatsu Photonics), lens-coupled to a high-speed scientific complementary metal oxide semiconductor sensor camera (C1440-22CU; Hamamatsu), 33 ms per frame. For fluorescence labelling of Halo7-tagged proteins, cells were incubated with 30nM STELLA 650-conjugated HaloTag ligand (GORYO) in Ham's F12 media (Invitrogen), 37˚C 20 min, washed x3, and media replaced by Ham's F12 media with 2 mM PIPES, pH7.0.

Tracking. For SW620:EGFR-GFP MATLAB (MathWorks) code (38) was used to track foci in green/red channels to determine spatial localization and calculate integrated pixel intensities and diffusion coefficients. The centroid of each focus was determined using iterative Gaussian masking to sub-pixel precision of 40nm, brightness calculated as the summed intensity inside a 5-pixel-radius centroid-centered circle, after subtraction of local background, signal-to-noise ratio (SNR) defined as intensity divided by background standard deviation. For SNR >0.3 (optimum for high true and low false positive detection from simulations trained on *in vitro* data) a focus was accepted and fitted with a two-dimensional (2D) radial Gaussian to determine its sigma

width. Foci detected in consecutive images separated by ≤5 pixels and not different in brightness or width by more than a factor of two were linked into the same track. For CHO-K1 foci tracking used a similar algorithm.

Stoichiometry. Stoichiometry per track was estimated in MATLAB using step-wise fluorophore photobleaching to determine GFP or TMR brightness (35) from live cells and corroborated *in vitro*. Live cell foci brightness followed exponential photobleaching. As each focus photobleaches it will emit the characteristic single GFP or TMR brightness value, I_{GFP} or I_{TMR}, detected as the peak of foci intensities over time. Estimates for I_{GFP} and I_{TMR} were verified by Fourier spectral analysis (35) yielding the same value within error. Initial intensity I_0 was estimated by interpolation of the first 3 points in each track, stoichiometries by dividing I_0 by the single-molecule fluorophore brightness, distributions rendered as kernel density estimations (35). Previous live-cell measurements using the same fluorescent protein (GFP) using photobleaching of all GFP followed by suppression of further GFP expression and then subsequent measurement of any reappearance of fluorescence intensity over a timescale up to 60 min have indicated that the typical proportion of immature GFP is <15% of the total (68), so no correction to stoichiometry was made.

Total Copy Number. The number of EGFR-GFP for SW620:EGFR-GFP on the cell surface was estimated by integration (37) of pixel intensities of the cell area in TIRF corrected for auto-fluorescence using parental SW620 strain. SW620:EGFR-GFP mean pixel fluorescence was calculated for every cell from a region segmented from brightfield using Sobel edge detection, morphologically dilating by a 7-pixel-radius disk to minimize cell-edge effects. Copy number was estimated by multiplying this value over the area of the cell, approximated as a 14 μm-diameter sphere.

Mobility. 2D mean square displacements (MSDs) of tracked clusters were calculated in MATLAB using the centroid positions at time t, $(x(t),y(t))$, considering N consecutive image frames with time interval $\tau=n\Delta t$, where n is a positive integer and Δt is the frame integration time (71):

$$MSD(\tau)=MSD(n\Delta t)=\frac{1}{N-1-n}\sum_{i=1}^{N-1-n}\left\{\left[x(i\Delta t+n\Delta t)-x(i\Delta t)\right]^2+\left[y(i\Delta t+n\Delta t)-y(i\Delta t)\right]^2\right\}$$

$$=4D\tau+4\sigma^2$$

The localization precision is σ=(40±20) nm. The apparent diffusion coefficient, D, was estimated from linear fits to the first three points in MSD *vs.* τ, constrained to pass through $4\sigma^2$ at $\tau = 0$ and allowing σ to vary in the range 20-60 nm.

Colocalization. For SW620:EGFR-GFP cells, colocalization between EGFR-GFP and EGF-TMR was calculated as the overlap integral between green/red foci in MATLAB, whose centroids were within 5 pixels (38). Assuming two 2D normalized Gaussian intensity distributions $g_1(x,y)$ and $g_2(x,y)$, centered around (x_1, y_1) with width σ_1 and centered around (x_2, y_2) with width σ_2 for green and red foci, respectively, the overlap integral v is (38):

$$v = exp\left(\frac{-\Delta r^2}{2(\sigma_1^2+\sigma_2^2)}\right),$$

where:

$\Delta r^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$.

Our simulations indicated that green/red foci pairs with identical centroids have an overlap of ~0.75, therefore we used a 0.75 threshold for colocalization detection.

For dwell time analysis in CHO-K1 we used a similar method (72, 73), detecting colocalized green/red foci pairs then measuring the duration over which this separation remained ≤180 nm, and generated a histogram distribution of these (denoted "0° image rotation dwell time distribution") that could be fitted by the sum of two exponential decay functions with time constants t_1=53±2 ms and t_2=335±100 ms. To determine the time constant associated with random colocalization we rotated the red images by 180° prior to performing the same analysis, indicating

a dwell time distribution (denoted "180° image rotation dwell time distribution") fitted with a single exponential function t_{rot}=52±2 ms, identical within error to t_1, suggesting that t_2 is associated with non-random colocalization. We tested the difference between the "0° image rotation dwell time distribution" and the "180° image rotation dwell time distribution" using a Brunner-Munzel test that indicated very significant differences ($p=3.1x10^{-5}$).

Modeling foci overlap probability. To assess whether the observed large EGFR clusters could be due to chance colocalization of small clusters combined with our optical resolution limit, we simulated cluster stoichiometry, using experimental fluorescent foci surface densities. Foci were simulated as 2D Gaussians with the same characteristic intensity as *in vivo,* and we used the same noise conditions as measured from real cell images. The probability that ≥2 fluorescent foci are separated by less than the optical resolution was determined in MATLAB using a previous model (38). In this model, given a surface density of spots ρ, the probability for a nearest neighbor spot to be within a distance w is given by:

$$p(w) = 1\text{-}\exp(-\pi w^2 \rho),$$

where, for overlapping foci, w is the optical resolution. The resulting probability of overlap can be used to generate the distribution of number of overlaps as a Poisson distribution. The apparent stoichiometry distribution from overlapping foci was then modeled by convolving this Poisson distribution with the expected intensity distribution of an isolated multimer, given by the intensity distribution of a single GFP (Figure S5) width scaled as the square root of the multimer stoichiometry. The latter was obtained by scaling the width of the single fluorophore intensity distribution (Figure S7) by $S^{1/2}$, where S is the model stoichiometry. A random Monte Carlo stoichiometry was generated from several simulated underlying EGFR molecular stoichiometries including monomers, dimers and the mixed oligomer model from Needham *et al.* (39) (51.3% monomer, 21.3% dimer, 10.3% trimer, 5.3% tetramer and 4.1% pentamer). Finally, a population distribution of oligomeric EGFR with stoichiometry sampled from a Poisson distribution with means varying from 1-10 but with broad distributions up to tens of molecules.

Statistics. Two-tailed Student's *t*-tests were performed for comparisons between pairs of datasets to test null hypothesis that data in each was sampled from the same statistical distribution assuming (n_1+n_2-2) degrees of freedom where n_1 and n_2 are the number of data points in each distribution and by convention that *t* statistic values which have a probability of confidence $p<0.05$ are statistically not significant. For TIRF each cell was defined as a biological replicate sampled from the cell population with sample sizes of 10-117 cells per condition. Technical replicates are not possible with irreversible photobleaching, nevertheless. Pairwise differences between stoichiometry distributions, and of colocalization dwell times, were also assessed using the non-parametric Brunner-Munzel rank order test, for which we assumed that $p<0.05$ indicates statistical significance. Explicit p values are reported where $p>0.0001$ and not reported below this threshold as they are no longer physically meaningful.

RT-qPCR. To extract RNA, cell pellets were lysed in Trizol (Invitrogen). RNA was converted into cDNA using MMLV reverse transcriptase (New England Biolabs®) with Oligo(dT)12-18 primers (Invitrogen), 10mM dNTP mix and RNase inhibitor Ribolock (Thermo Fisher Scientific), cDNA purified using QIAquick PCR purification (QIAGEN). Expression levels of *HER2*, *HER3*, *HER4* and *EGFR* were determined by qPCR using Fast SYBR Green Master Mix on QuantStudio TM 3 Real-Time PCR System (Thermo Fisher Scientific), 20s/95°C then 40 cycles of 1s/95°C and 20s/60°C, normalized against housekeeping *PLQC2*. Relative fold expression change was calculated using ΔΔCt analysis.

Microarray.

Microarray expression data were generated by a service provided by the Patterson Laboratory in Manchester, UK. The gene expression data for 78 unique, non-duplicate (not sourced from same patients) colon cancer cell lines were obtained by performing microarray using the Affymetrix GeneChip HG-U133 Plus 2.0 microarray. Data is normalized using RMA and batch-removed using Partek Genomics Suite software. These were compared against publicly available RNA-Seq

data (74). For microarray, we developed an algorithmic pipeline with the goal of identifying subpopulations within a given data through a statistical machine-learning method. With this pipeline we statistically determined the expressing and non-expressing subpopulations in our 78 CRC microarray cell line data. The statistically determined background threshold closely matches with the background expression range of 100-150 microarray counts shown in previous work done by our lab; this therefore defines our background threshold. Expression level below the background threshold was considered as non-expressing.

The results obtained from different probesets for the microarray where available, and for RNAseq, are given below for both the SW620 cell line. In the case of the TGFA ligand we also compared the SW620 output with that of SW480 (a duplicate cell line of SW620 derived from the same patient):

TGFA ligand:

	SW620	**SW480**	**Category**
Microrarray 211258_s_at	100	81	Background
Microrarray 205015_s_at	218	195	Very Low
Microrarray 205016_at	360	327	High
CCLE RNA-Seq (RPKM)	6.04	11.04	Very Low

(**211258_s_at and 205015_s_at**)
Pearson correlation: 0.93
P-value: 4.27e-35
Note: 2 of the 3 probesets of TGFA (211258_s_at and 205015_s_at) correlate strongly with one another with a r-value of 0.93 while either probesets correlate with 205016_at with a r-value of 0.74 and 0.76, respectively. Further exploring SW480, a duplicate cell line of SW620 derived from the same patient, revealed a similar pattern in all three probes. We conclude that 205106_at is an anomaly that does not follow the pattern of the other two probes. Adding onto the expression level observed in CCLE RNA-Seq data of 6.04, we can conclude a consensus very low expression level of TGFA.

ERBB2 ligand:

	SW620	**Category**
Microrarray 216836_s_at	200	Very Low
Microrarray 210930_s_at	74	Background
CCLE RNA-Seq (RPKM)	9.07	Very Low

(**216836_s_at and 210930_s_at**)
Pearson correlation: 0.50
P-value: 2.53e-06
Note: Both probesets of ERBB2 (216836_s_at and 210930_s_at) do not correlate as well with a r-value of 0.5 thus we score them individually and look at them separately. Adding onto the expression level observed in CCLE RNA-Seq data, we can conclude a consensus low expression level of ERBB2.

HBEGF ligand:

	SW620	**Category**
Microrarray 222076_at	41	Background
Microrarray 244857_at	87	Background
Microrarray 38037_at	189	Very Low
Microrarray 203821_at	239	Very Low
CCLE RNA-Seq (RPKM)	1.53	Very Low

(**222076_at and 203821_at**)
Pearson correlation: 0.97
P-value: 5.81e-52
Note: 2 of the 4 probesets of HBEGF (222076_at and 203821_at) correlate strongly with an r-value of 0.97 while 38037_at has a poor correlation with either probesets of 0.02 and 0.02, respectively, and 203821_at has a poor correlation with either probesets of 0.11 and 0.13, respectively. Surprisingly, 38037_at and 203821_at also have a poor correlation with an r-value of 0.07, suggesting that the two probesets have unique patterns. We conclude that both 38037_at and 203821_ are anomalies that do not behave in line with the other two probes. Thus, 222076_at and 203821_at is assumed to represent the general expression pattern of HBEGF. Adding onto the expression level observed in CCLE RNA-Seq data, we conclude a consensus low expression level of ERBB2.

AREG ligand:

	SW620	**Category**
Microrarray 205239_at	97	Background
Microrarray 215564_at	34	Background
CCLE RN-Seq (RPKM)	0.103	Background

(**205239_at and 215564_at**)
Pearson correlation: 0.17
P-value: 0.13
Note: Both probesets of AREG (205239_at and 215564_at) do not correlate well with an r-value of 0.17 thus we score them individually and look at them separately. Adding onto the expression level observed in CCLE RNA-Seq data, we can conclude a consensus background expression level of AREG.

EREG ligand:

	SW620	**Category**
Microrarray 1569583_at	59	Background
Microrarray 205767_at	133	Background
CCLE RNA-Seq (RPKM)	0.407	Background

(**1569583_at and 205767_at**)
Pearson correlation: 0.81
P-value: 3.77e-19
Discussion: Both probesets of EREG (1569583_at and 205767_at) correlate strongly with an r-value of 0.81. Adding onto the expression level observed in CCLE RNA-Seq data, we can conclude a consensus background expression level of EREG.

BTC ligand:

	SW620	**Category**
Microrarray 207326_at	39	Background
Microrarray 241412_at	16	Background
CCLE RNA-Seq (RPKM)	0.127	Background

(**207326_at and 241412_at**)
Pearson correlation: 0.87
P-value: 1.50e-25
Note: Both probesets of BTC (207326_at and 241412_at) correlate strongly with an r-value of 0.87. Adding onto the expression level observed in CCLE RNA-Seq data, we conclude a consensus background expression level of BTC.

EGF ligand:

	SW620	Category
Microarray 206254_at	26	Background
CCLE RNA-Seq (RPKM)	0.0029	Background

Discussion: EGF has a single probeset with a background expression level. Adding onto the expression level observed in CCLE RNA-Seq data, we can conclude a consensus background expression level of EGF.

EPGN ligand:

	SW620	Category
CCLE RNA-Seq (RPKM)	0	Background

Note, there is no available microarray data for EPGN. The CCLE RNA-Seq data showed a level of 0 thus we can conclude a background expression level of EPGN.

Lastly, a good example of a high expression level in CRC cell lines is the marker CDH1 since it is the signature gene expressed by all epithelial cells. In our in-house microarray data, the average expression level of CDH1 is ~6000 while in the CCLE RNA-Seq data lie around 45-60.

Kinetic modeling of EGFR ligand binding, dimerization, endocytosis and recycling

We developed a multi-state time-dependent kinetics model for ligand binding to receptor monomers and dimers, incorporating homo- and hetero-dimerization of ligated and unligated receptors, internalization of ligated receptors via endocytosis and subsequent recycling of receptors to the plasma membrane. The model solves multiple rate equations to determine concentrations of ligated and unligated receptor monomers and dimers, and concentrations of internalized receptors, as a function of time (Supplementary). To match our experiments, the rate constants for ligand binding and unbinding used are based on previous reports for receptor monomers at 37°C (equilibration time scale ~5-10 min) (75, 76). Unligated dimers have the same ligand-binding rate as unligated monomers, whereas singly-ligated dimers have a 14-fold lower binding rate than unligated monomers, as reported from equilibrium measurements at 4°C (8) (details in Supplementary). We do not explicitly model oligomeric or clustered EGFR states, as their rates are unknown. The model assumes the presence of populations of EGFR monomers and dimers (Figure 4a) with higher-order clusters formed by coalesce of monomers and dimers in the plasma membrane. Receptor dimerization rates were chosen to match the relatively fast typical time scales of equilibration of dimer formation in living cells at 37°C (~0.1-1s, see (77)) as well as the fact that we observe few receptor monomers on the cell surface at equilibrium. The unligated-unligated and unligated-ligated receptor dimerization rates are the same, whereas ligated-ligated dimerisation has a 10-fold lower rate as per equilibrium observations at 4°C (8) (details in Supplementary). Under the conditions of our experiments of relatively high EGF concentration where we likely saturate EGFRs at the surface, the rate of internalization is 3-10%/min, dependent on cell line (8, 76). This is lower than that at lower EGF concentrations owing to clathrin endocytosis pathway saturation. Recycling rates of ligand-occupied EGFR are ~10%/min (8), with recycling contributing significantly to the overall receptor distribution only after a pool of endosomal EGFR is accumulated.

The time-dependent kinetic model that we have developed considers: (i) that EGF ligand can bind to both EGFR receptor monomers and dimers, (ii) that ligated EGFR monomers and singly- or doubly-ligated dimers can be internalized via endocytosis, and (iii) that endocytosed receptors can be recycled back to the plasma membrane. EGFR degradation is not included in the model, as the half-life of EGFR degradation after EGF activation is typically of the order of hours(48) and hence has a negligible impact considering the time scale of our measurements (~40 min).

In what follows, we use the following notation: R for receptor monomers, RR for receptor dimers, L for EGF ligand, RL for ligated receptor monomers, RRL for singly-ligated receptor dimers, RRL2 for doubly-ligated receptor dimers, RL^{inside} for endocytosed RL, RRL^{inside} for endocytosed RRL and $RRL2^{inside}$ for endocytosed RRL2. Our model considers the following reversible reactions for ligand binding and receptor dimerization:

$R + L \rightarrow RL$ with on-rate constant $k_{11,on}$, off-rate constant $k_{11,off}$ and equilibrium association constant $K_{11} = k_{11,on}/k_{11,off}$;

$RR + L \rightarrow RRL$ with on-rate constant $k_{21,on}$, off-rate constant $k_{21,off}$ and equilibrium association constant $K_{21} = k_{21,on}/k_{21,off}$;

$RRL + L \rightarrow RRL2$ with on-rate constant $k_{22,on}$, off-rate constant $k_{22,off}$ and equilibrium association constant $K_{22} = k_{22,on}/k_{22,off}$;

$R + R \rightarrow RR$ with on-rate constant $l_{20,on}$, off-rate constant $l_{20,off}$ and equilibrium dimerization constant $L_{20} = l_{20,on}/l_{20,off}$;

$RL + R \rightarrow RRL$ with on-rate constant $l_{21,on}$, off-rate constant $l_{21,off}$ and equilibrium dimerization constant $L_{21} = l_{21,on}/l_{21,off}$,;

$RL + RL \rightarrow RRL2$ with on-rate constant $l_{22,on}$, off-rate constant $l_{22,off}$ and equilibrium dimerization constant $L_{22} = l_{22,on}/l_{22,off}$.

Additionally, we have the following non-reversible endocytosis and recycling reactions:

$RL \rightarrow RL^{inside} \rightarrow R$

$RRL \rightarrow RRL^{inside} \rightarrow RR$
$RRL2 \rightarrow RRL2^{inside} \rightarrow RR$.

For the last three reactions, the first arrow (endocytosis) is considered to have a rate k_{endoc} and the second arrow (recycling back to the plasma membrane) is considered to have a rate $k_{recycle}$. We consider only endocytosis of ligated receptors (RL, RRL, RRL2) and, in the first instance, we assume that recycling returns unligated receptors to the plasma membrane. We assume the same endocytosis and recycling rates for the three reactions above.
Considering all the above reactions, it is possible to write a system of simultaneous rate equations as follows:

$$\frac{d[RR]}{dt} = l_{20,on}[R]^2 - l_{20,off}[RR] - k_{21,on}[RR][L] + k_{21,off}[RRL] + k_{recycle}[RRL^{inside}] + k_{recycle}[RRL2^{inside}],$$

$$\frac{d[RL]}{dt} = k_{11,on}[R][L] - k_{11,off}[RL] - k_{endoc}[RL] - l_{21,on}[RL][R] + l_{21,off}[RRL] - l_{22,on}[RL]^2 + l_{22,off}[RRL2],$$

$$\frac{d[RRL]}{dt} = k_{21,on}[RR][L] - k_{21,off}[RRL] - k_{endoc}[RRL] + l_{21,on}[RL][R] - l_{21,off}[RRL] - k_{22,on}[RRL][L] + k_{22,off}[RRL2],$$

$$\frac{d[RRL2]}{dt} = k_{22,on}[RRL][L] - k_{22,off}[RRL2] - k_{endoc}[RRL2] + l_{22,on}[RL]^2 - l_{22,off}[RRL2],$$

$$\frac{d[RL^{inside}]}{dt} = k_{endoc}[RL] - k_{recycle}[RL^{inside}],$$

$$\frac{d[RRL^{inside}]}{dt} = k_{endoc}[RRL] - k_{recycle}[RRL^{inside}],$$

$$\frac{d[RRL2^{inside}]}{dt} = k_{endoc}[RRL2] - k_{recycle}[RRL2^{inside}],$$

$$[R] = R_{total} - ([RL] + [RR] + [RRL] + [RRL2] + [RL^{inside}] + [RRL^{inside}] + [RRL2^{inside}]).$$

Here *t* is time and the square brackets indicate concentration of reactant/product, with all concentrations being functions of time, and all rates defined as above. R_{total} is the total concentration of receptors (including both receptors on the surface and internalised), i.e., the total number of receptor molecules per cell. This value is set to 200,000-400,000 molecules/cell (depending on the model parameters used) to match our observations of approximately 200,000 receptors on the cell surface in experiments.

In order to solve the system of equations we consider the following initial conditions: $[RL](t=0) = [RRL](t=0) = [RRL2](t=0) = [RL^{inside}](t=0) = [RRL^{inside}](t=0) = [RRL2^{inside}](t=0) = 0$, i.e., that all ligated and internalised components are zero at t=0, and we set the initial concentrations of unligated receptor monomers and dimers to certain fractions of R_{total}, typically ~5% monomer fraction and ~95% dimer fraction, in agreement with the fact that we observe mostly clusters of receptors and hardly any monomers in our experiments. The ligand concentration, [L], used is that in our experiments: 100 ng/ml ~ 15.6 nM.

The rate constants for monomer ligand binding/unbinding are given values $k_{11,on} = 10^6 M^{-1}s^{-1}$ and $k_{11,off} = 10^{-3}s^{-1}$, so that the equilibrium association constant is $K_{11} = 10^9 M^{-1}$. These rate constants are similar to those previously reported for EGFR monomers(76) and ensure that the time scale of equilibration of receptor binding is ~5-10 min, in agreement with measurements reported in live cells at 37°C (75, 76). We assume that the ligand binding/unbinding rate constants for unligated receptor dimers are the same as for unligated monomers, i.e., $k_{21,on} = k_{11,on}$ and $k_{21,off} = k_{11,off}$ (hence $K_{21} = K_{11}$). This assumption is motivated by the fact that the equilibrium association constants for R and RR binding to ligand for EGFR have been reported to be approximately the same, i.e., $K_{21} \approx K_{11}$, as obtained from measurements in cells at 4°C(8). The same study reported that the equilibrium association constant for ligand binding to singly-ligated receptor dimers is ~14 times lower than that for unligated monomers/dimers, i.e., $K_{22} \approx K_{11}/14$. To include this reported lower affinity of ligand for singly-ligated receptors, we choose on/off-rate constants for ligand binding to RRL of $k_{22,on} = k_{11,on}/7$ and $k_{22,off} = 2 \times k_{11,off}$. Note that these conditions imply that there is negative cooperativity of EGFRs in ligand binding, as $K_{22} < K_{11}/4$.

As for receptor dimerization, equilibration is achieved in shorter times, of the order of ~0.1-1s, as measured for G-protein-coupled receptors (GPCRs) in living cells at 37°C (77). For our model, we choose an equilibrium dimerization constant $L_{20} = 10^{-3}(molecules/cell)^{-1}$ in units of 2D concentration (with 1 receptor molecule per μm^2 corresponding to approximately 3,000 receptors/cell), 10 times higher than the value measured for GPCRs. We use $l_{20,on} = 10^{-4}(molecules/cell)^{-1}s^{-1}$ and $l_{20,off} = 10^{-1}s^{-1}$ to guarantee that the time scale of dimer formation equilibration is correct and to ensure that there are very few monomers on the cell surface at equilibrium (~5% monomers, ~95% dimers), in agreement with our own experimental observations. For the dimerization rate of RL and R, we use $l_{21,on} = l_{20,on}$ and $l_{21,off} = l_{20,off}$ as the respective equilibrium dimerization constants reported from measurements in cells at 4°C are very similar ($L_{21} \approx L_{20}$) (8). As the dimerization rate L_{22} for RL+RL was reported to be ~10 times lower than L_{20} in the same study, we choose as model parameters $l_{22,on} = l_{20,on}/5$ and $l_{22,off} = 2 \times l_{20,off}$.

We set the rate of endocytosis of ligated receptors in our model to $k_{endoc} = 6\,\%/min = 0.06/60\,s^{-1}$, in agreement with previous reports of EGFR endocytosis rates in living cells at 37°C in the presence of high EGF concentrations similar to that in our experiments, which are in the range $3 - 10\,\%/min$ (8, 76). EGFR endocytosis rates at low ligand concentrations (near physiological levels [EGF]~1ng/ml) are typically in the range $15 - 30\,\%/min$. At higher EGF concentrations (~100 ng/ml) and high receptor numbers (as in our experiments), the clathrin-endocytosis pathway saturates, and the endocytosis rate is lower ($3 - 10\,\%/min$). Endocytosis rates in the absence of EGF ligand are greatly decelerated with respect to the above rates(8). The recycling rate is set to $k_{recycle} = 10\,\%/min = 0.1/60\,s^{-1}$ (8).

Using all the above parameter values, we solve the above system of differential equations in order to find [R], [RL], [RR], [RRL], [RRL2], [RL^{inside}], [RRL^{inside}] and [RRL2^{inside}] as a function of time. Additionally, we calculate the fractional saturation on the cell surface, $Y^{surface}$, equal to the ratio of EGF to EGFR molecules (EGF:EGFR ratio) on the plasma membrane. This is given by:

$$Y^{surface} = \frac{[R_{bound}]^{surface}}{[R_{bound}]^{surface}+[R_{unbound}]^{surface}} = \frac{[RL]+[RRL]/2+[RRL2]}{[R]+[RL]+[RR]+[RRL]+[RRL2]}\ .$$

Here, we account for the fact that all concentrations involving receptors in the model are obtained as numbers of receptor molecules per cell. The value of the EGFR:EGF ratio obtained from our experimental measurements is ~4 on average (and ~2 as a modal peak value), which would correspond to a $Y^{surface}$ value in our model of ~0.25 (and peak value ~0.5).

The calculated variables as a function of time for the above-mentioned parameters are shown in Figure 4a in the main text. We use R_{total} = 310,000 molecules/cell, which is the value that results in a total number of receptor molecules on the cell surface (excluding internalised receptors) at equilibrium of 200,000, in agreement with our observations. We find that the fractional saturation at the cell surface is $Y^{surface} \sim 0.66$, i.e., the EGFR:EGF ratio predicted by the model is ~1.5.

The results for $Y^{surface}$ (at equilibrium) do not change significantly upon various changes of the model parameters. Changing the value of R_{total} by ±100,000 receptors/cell only alters the result by <2%. Increasing the endocytosis rate to 10%/min yields $Y^{surface}0.62$ and decreasing it to 3%/min yields $Y^{surface}0.72$. Increasing the recycling rate up to 40%/min does not change the value of $Y^{surface}$ at equilibrium, decreasing it to 0 results in $Y^{surface}0.83$ and a largely reduced number of receptors at the cell surface (~30,000) after 40 minutes. Increasing or decreasing all ligand binding on-rate constants by a factor of 2 results in moderately increased or decreased values at equilibrium, $Y^{surface}0.77$ or $Y^{surface}0.55$, respectively. Increasing or decreasing all ligand binding off-rate constants by a factor of 2 results in moderately decreased or increased values, $Y^{surface}0.61$ or $Y^{surface}0.70$, respectively. Increasing or decreasing all receptor dimerization on-rate constants by a factor of 2 changes the result only by <3%. Changing the factors in $k_{22,on}$ and $k_{22,off}$ that guarantee that $K_{22} = K_{11}/14$, to go from $k_{22,on} = k_{11,on}/7$ and $k_{22,off} = 2 \times k_{11,off}$ to the extremes $k_{22,on} = k_{11,on}/14$, $k_{22,off} = k_{11,off}$ or $k_{22,on} = k_{11,on}$, $k_{22,off} = 14 \times k_{11,off}$ changes the result by at most 3%. Changing the factors that

guarantee that $L_{22} = L_{20}/10$ to the extremes $l_{22,on} = l_{20,on}/10,\ l_{22,off} = l_{20,off}$ and $l_{22,on} = l_{20,on},\ l_{22,off} = 10 \times l_{20,off}$, does not change the result at all. Reducing all dimerization on/off-rate constants by a factor of 100 (which increases the dimerization equilibration time from ~1s to ~1min) only increases $Y^{surface}$ to 0.68 (by 3%). Changing the initial fractions of receptor monomers and dimers at t=0 does not change the results as the equilibration time of dimerization is short (~1s).

Setting the binding on-rate constant for singly-ligated dimers to zero ($k_{22,on} = 0$), i.e. assuming extreme negative cooperativity, the result remains the same ($Y^{surface}0.66$). Setting the binding on-rate constant for unligated dimers to zero ($k_{21,on} = 0$) makes a significant difference and results in $Y^{surface}0.24$, i.e., in an EGFR:EGF ratio at equilibrium on the cell surface of ~4 (see Figure 4b in the main text). This is the case regardless of whether $k_{22,on}$is also set to zero or not, i.e., this result is consistent also with extreme preferential ligand binding to monomers (ligand can bind to receptor monomers but not to dimers). Setting the binding on-rate constant for receptor monomers to zero ($k_{11,on} = 0$), i.e., assuming extreme preferential ligand binding to dimers, yields $Y^{surface}0.63$. In the case of positive cooperativity, setting $k_{22,on} = k_{11,on}$ and $k_{22,off} = k_{11,off}$, the result increases to $Y^{surface}0.69$. Considering more significant positive cooperativity with $k_{22,on} = 10k_{11,on}$ and $k_{22,off} = k_{11,off}$, we obtain an even higher value, $Y^{surface}0.82$. Hence, positive cooperativity leads to higher values of $Y^{surface}$, while negative cooperativity leads to lower values. Hence, the latter agrees better with our experimental results that correspond to an EGFR:EGF ratio ~4 on average (i.e., to $Y^{surface}\ 1/4$). Model predictions in the case of preferential ligand binding to monomers would agree with experimental observations. We also tried setting some of the dimerization rate constants to zero as follows. Assuming that R monomers do not dimerize to form RR dimers ($l_{20,on} = 0$), we obtain $Y^{surface}0.49$ and $Y^{surface}0.69$ if we also set the off-rate constant to zero ($l_{20,off} = 0$). Assuming that RL and R do not dimerize to form RRL ($l_{21,on} = 0$), we obtain $Y^{surface}0.92$ and $Y^{surface}0.69$ if we also set the off-rate constant to zero ($l_{21,off} = 0$). Hence, a decreased hetero-dimerization rate constant (ligated-unligated) leads to model predictions that are further from our experimental observations. Assuming that RL complexes do not dimerize to form RRL2 ($l_{22,on} = 0$), we obtain $Y^{surface}0.53$ and $Y^{surface}0.68$ if we also set the off-rate to zero ($l_{22,off} = 0$). Setting both R and RL homo-dimerization rate constants (unligated-unligated and ligated-ligated) to zero ($l_{20,on} = 0$ and $l_{22,on} = 0$), the model yields $Y^{surface}0.47$. Hence, a reduced homo-dimerization on-rate constant for unligated and/or ligated monomers (i.e., a reduced $l_{20,on}$ and/or a reduced $l_{22,on}$) yields model results that are closer to our experimental observations (a lower $Y^{surface}$ value).

We also tested in our model ligand binding and dimerization rate constants that result in the same equilibrium association constants as those reported from measurements in cells at 4°C, i.e., $K_{11} \approx 5 \times 10^9 M^{-1}$, $K_{21} \approx K_{11}$, $K_{22} \approx K_{11}/14$, $L_{20} \approx 2 \times 10^{-5}(molecs/cell)^{-1}$, $L_{21} \approx L_{20}$ and $L_{22} \approx L_{20}/10$ (8). In this case, we set the model parameters to R_{total} = 310000 molecules/cell, 5% and 95% initial fractions of monomers and dimers respectively, $[L] = 15.6 \times 10^{-9} M$, $k_{endoc} = 0.06/60\ s^{-1}$, $k_{recycle} = 0.1/60\ s^{-1}$ (up to here all parameters are the same as previously), and $k_{11,on} = 10^7 M^{-1}s^{-1}$, $k_{11,off} = 2 \times 10^{-3}s^{-1}$, $k_{21,on} = k_{11,on}$, $k_{21,off} = k_{11,off}$, $k_{22,on} = k_{11,on}/7$ and $k_{22,off} = 2 \times k_{11,off}$, $l_{20,on} \approx 2 \times 10^{-5}(molecs/cell)^{-1}s^{-1}$ and $l_{20,off} = 1s^{-1}$, $l_{21,on} = l_{20,on}$, $l_{21,off} = l_{20,off}$, $l_{22,on} = l_{20,on}/5$ and $l_{22,off} = 2 \times l_{20,off}$. Using these parameters, the value of the surface fractional saturation (EGF:EGFR ratio) is $Y^{surface}0.96$ (see Figure 4c in the main text). The result remains the same if all binding on/off-rate constants are multiplied or divided by 10, or if all dimerization on/off-rate constants are multiplied or divided by 10, keeping the equilibrium association constants the same. This value of $Y^{surface}0.96$ close to 1 at our ligand concentration, as well as values obtained at ligand concentrations in the range 10^{-11}-10^{-6} M (Figure 4c, main text) are all consistent with results at 4°C from alternative models based on equilibrium equations (that do not include time

dependence) (8), proving the validity of our model. It is important to note that the on/off-rate constants can have a strong temperature dependence and that results measured at 4°C can differ significantly (up to a factor of 10-100) from measurements at 37°C (78).

In conclusion, the value we obtain experimentally for the EGFR:EGF ratio in living cells at 37°C agrees better with model predictions in the cases in which there is negative cooperativity for ligand binding and preferential ligand binding to monomers or reduced homo-dimerization on-rate constants (for unligated-unligated and ligated-ligated dimerization).

Additionally, we considered a model in which ligated receptors are recycled back to the plasma membrane ligated, as opposed to unligated. In this case, the system of rate equations that we solve is the following:

$\frac{d[RR]}{dt} = l_{20,on}[R]^2 - l_{20,off}[RR] - k_{21,on}[RR][L] + k_{21,off}[RRL]$,

$\frac{d[RL]}{dt} = k_{11,on}[R][L] - k_{11,off}[RL] - k_{endoc}[RL] - l_{21,on}[RL][R] + l_{21,off}[RRL] - l_{22,on}[RL]^2 + l_{22,off}[RRL2] + k_{recycle}[RL^{inside}]$,

$\frac{d[RRL]}{dt} = k_{21,on}[RR][L] - k_{21,off}[RRL] - k_{endoc}[RRL] + l_{21,on}[RL][R] - l_{21,off}[RRL] - k_{22,on}[RRL][L] + k_{22,off}[RRL2] + k_{recycle}[RRL^{inside}]$,

$\frac{d[RRL2]}{dt} = k_{22,on}[RRL][L] - k_{22,off}[RRL2] - k_{endoc}[RRL2] + l_{22,on}[RL]^2 - l_{22,off}[RRL2] + k_{recycle}[RRL2^{inside}]$,

$\frac{d[RL^{inside}]}{dt} = k_{endoc}[RL] - k_{recycle}[RL^{inside}]$,

$\frac{d[RRL^{inside}]}{dt} = k_{endoc}[RRL] - k_{recycle}[RRL^{inside}]$,

$\frac{d[RRL2^{inside}]}{dt} = k_{endoc}[RRL2] - k_{recycle}[RRL2^{inside}]$,

$[R] = R_{total} - ([RL] + [RR] + [RRL] + [RRL2] + [RL^{inside}] + [RRL^{inside}] + [RRL2^{inside}])$.

The solution we obtain with the original set of parameters (R_{total} = 310000 molecs/cell, 5% and 95% initial fractions of monomers and dimers respectively, $[L] = 15.6 \times 10^{-9} M$, $k_{endoc} = 0.06/60\, s^{-1}$, $k_{recycle} = 0.1/60\, s^{-1}$, $k_{11,on} = 10^6 M^{-1} s^{-1}$, $k_{11,off} = 10^{-3} s^{-1}$, $k_{21,on} = k_{11,on}$, $k_{21,off} = k_{11,off}$, $k_{22,on} = k_{11,on}/7$ and $k_{22,off} = 2 \times k_{11,off}$, $l_{20,on} \approx 10^{-4} (molecs/cell)^{-1} s^{-1}$, $l_{20,off} = 0.1 s^{-1}$, $l_{21,on} = l_{20,on}$, $l_{21,off} = l_{20,off}$, $l_{22,on} = l_{20,on}/5$ and $l_{22,off} = 2 \times l_{20,off}$) is $Y^{surface} 0.81$. Similarly, to what occurred for the previously mentioned model predictions for experiments at 37°C, the value of the surface fractional saturation decreases significantly when setting $k_{21,on}$ to zero, to $Y^{surface} 0.36$. Variations upon changes of parameters are very similar to those for the previous model at 37°C. Results from this second model are somewhat further from our experimental observations but the conclusions from above remain valid for both types of recycling processes considered (ligated receptors being endocytosed and recycled back to the plasma membrane unligated or ligated).

Considering our model predictions, we can account for our data by a combination of negative cooperativity of binding and decreased affinity of ligand for dimers. Reduced homo-dimerization on-rates might also play a role, as they result in lower $Y^{surface}$ values and hence in higher EGFR:EGF ratios, closer to our measured data (Figure 4d, Figure S10, Supplementary). These predictions could be consistent with initial EGF binding to monomeric EGFR to generate an activated state predisposed to dimerize with unligated EGFR. Figure 4e (solid green line) validates our model by showing that it can reproduce previously published (22) equilibrium results measured at 4°C as a function of ligand concentration.

Supplementary Figures

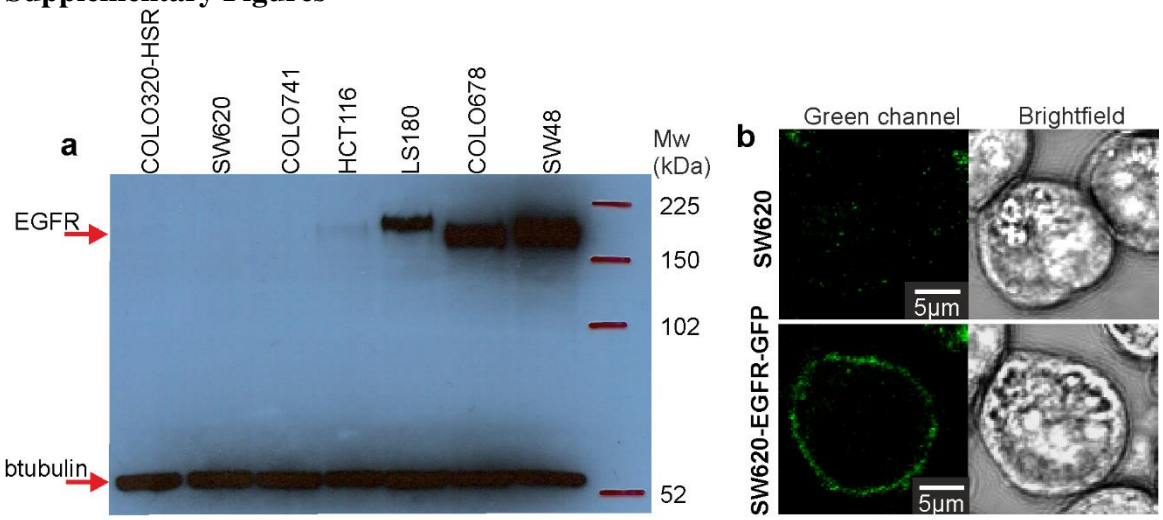


Figure S1: ***(a)*** *SDS-PAGE for candidate colon carcinoma cell lines, indicating that SW620 COLO320-HSR cells have negligible endogenous EGFR expression compared to positive controls HCT116, LS180, COLO678 and SW48 cells.* ***(b)*** *Parental SW620 shows minimal autofluorescence (upper left), while stably transfected SW620-EGFR-GFP (lower left) show plasma membrane localization of EGFR-GFP from confocal imaging of cells focusing at mid-cell-height (soon after adhering to the coverslip surface).*

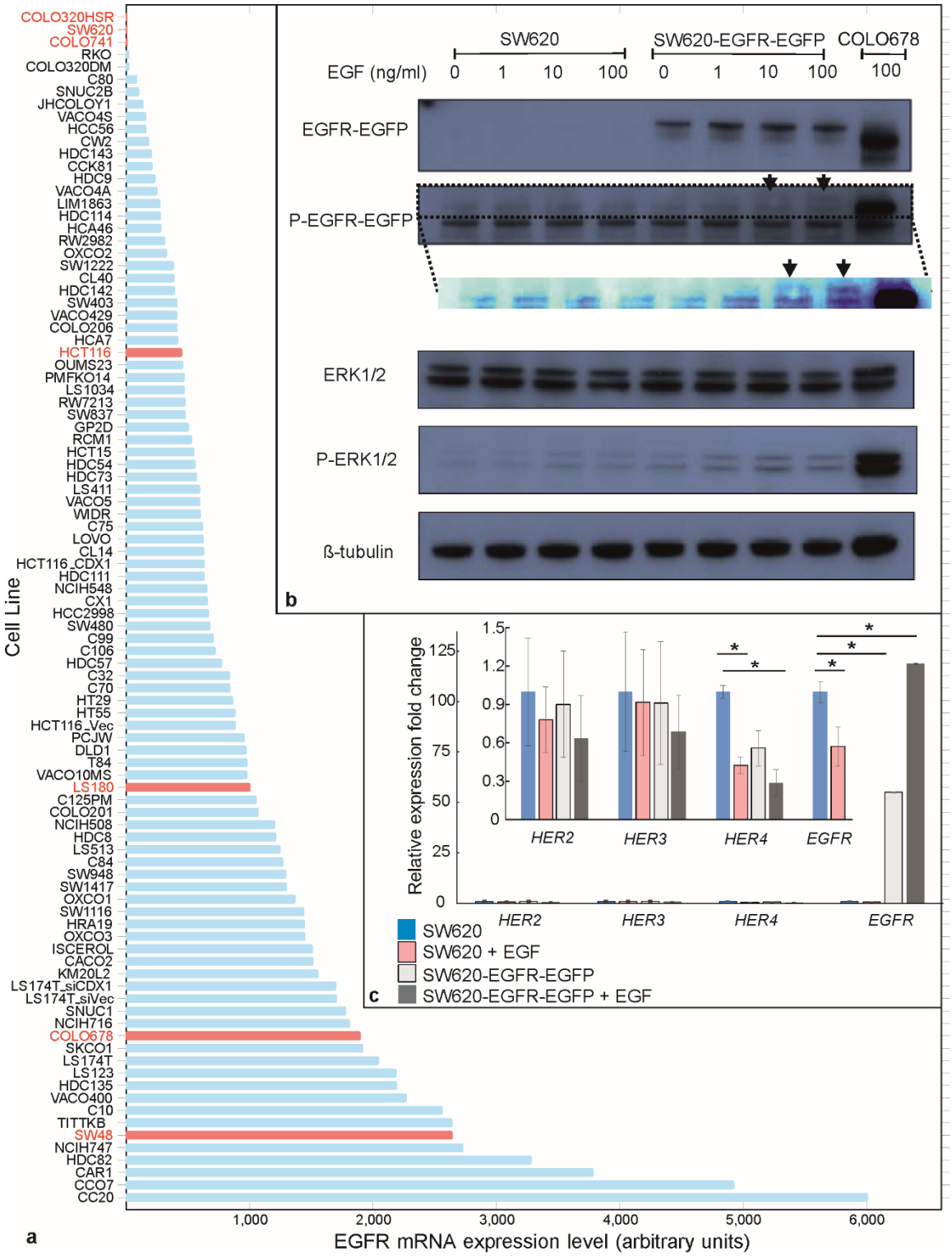


Figure S2: EGFR expression levels. (a). *The mRNA expression levels were quantified for a colon cancer-cell line panel using Affymetrix U133+2 mRNA microarray data. Measurements indicated three candidate cell lines, SW620, COLO320HSR and COLO741 (labeled in red, top of panel), as having very low or absent levels of native EGFR expression, as tested in subsequent western blot analysis in comparison to EGFR-expressing cell lines as positive controls (indicated as red columns, middle and bottom of panel). A further four positive controls with medium to high levels (HCT116, LS180, COLO678; indicated as red columns, middle and bottom of panel) can be seen. (**b**). SDS-PAGE performed on protein extracts obtained from SW620 wild type and SW620-EGFR-GFP cells before and after EGF treatment. COLO680 is used as a positive control. SW620 extracts do not have levels of EGFR detectable by SDS-PAGE, unlike SW620-EGFR-GFP, confirming successful transfection and expression of the GFP-tagged EGFR. EGF treatment of the transfected line results at concentrations of 10ng/ml or above results in detectable levels of phosphorylated EGFR (black arrows and inset with*

enhanced display contrast). EGF treatment of SW620-EGFR-GFP results in increased levels of phosphorylated ERK1/2 kinases, downstream players of the EGFR/ERK pathway, confirming the kinase activity of the receptor. ***(c).*** *Fold expression change of HER2, HER3, HER4 and EGFR relative to the PQLC2 house-keeping gene in SW620 and SW620-EGFR-EGFP cells with or without EGF treatment. The relative fold expression change was calculated based on the ΔΔCt method. The results represent the mean values and standard errors of four experiments on two biological replicates. *=Student t-test P<0.05.*

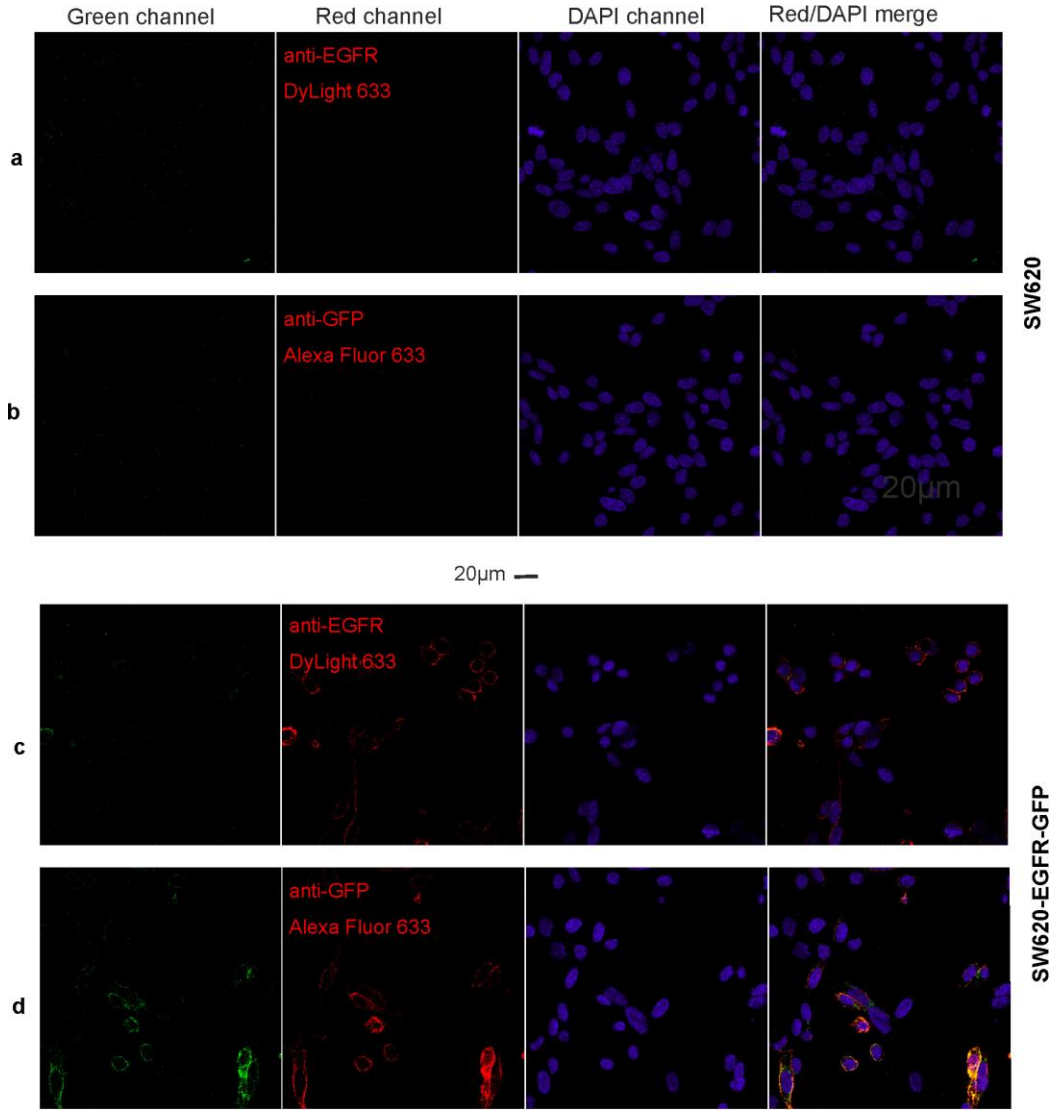


Figure S3: *Confocal microscopy images of fixed cells using GFP, anti-GFP immunofluorescence, and DAPI staining:* ***(a,b)*** *non-GFP background cell line SW620;* ***(c,d)*** *SW620-EGFR-GFP.*

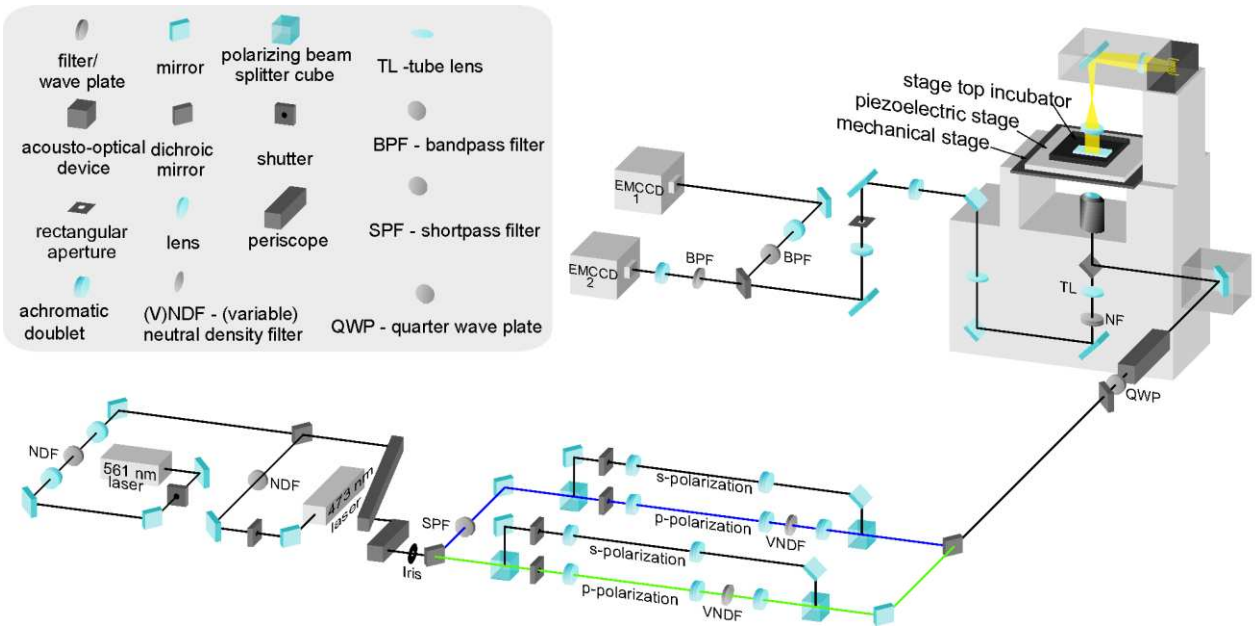


Figure S4: *optical-path diagram of bespoke single-molecule TIRF microscope.*

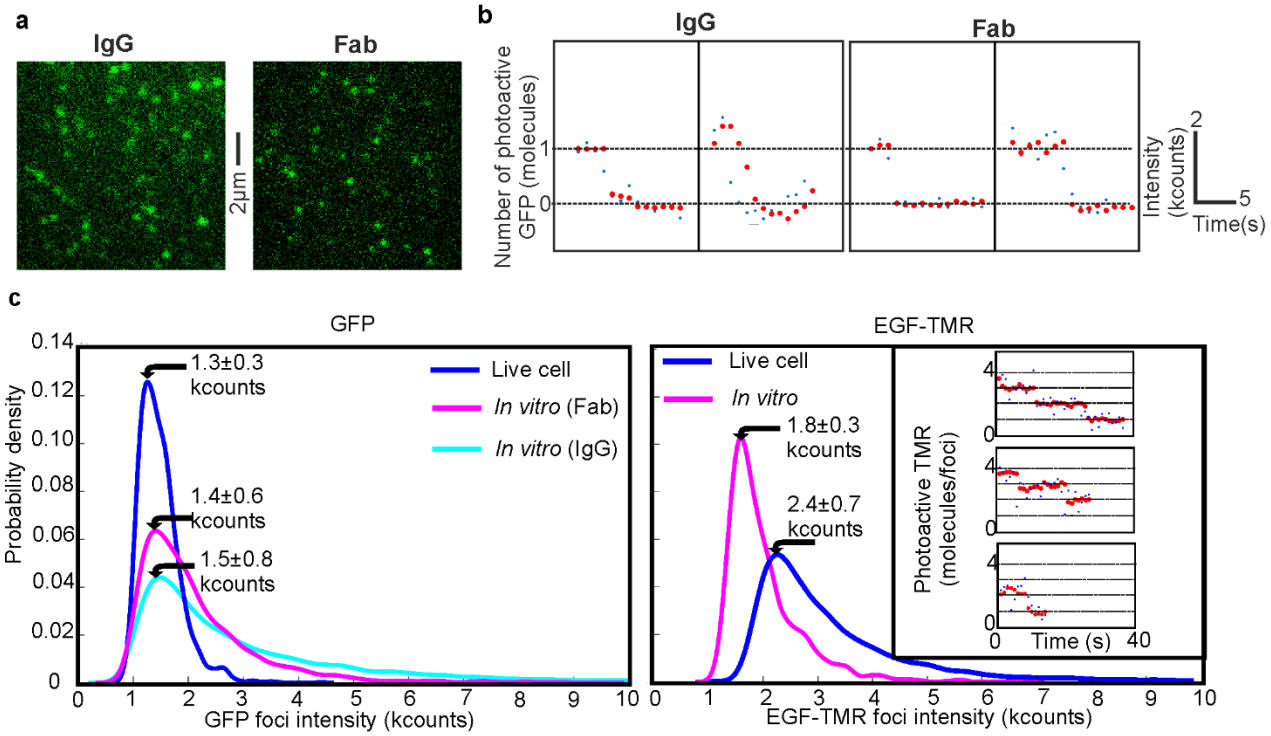


Figure S5: Characterization of unitary fluorophore brightness values. (a) *TIRF of GFP in vitro using IgG and Fab conjugation.* ***(b)*** *Step-wise photobleaching showing raw data (blue) and output data of an edge-preserving filter (red) given in kcounts (i.e. counts on detector times 10^3).* ***(c)*** *Kernel density estimation* (35) *distributions* (79, 80) *of fluorescent foci intensity values measured in kcounts.* ***Left panel****: measurements for single GFP in a live cell at the end of the photobleach, before EGF is added, are compared with in vitro Fab and whole IgG data. Arrows indicate mean and s.e.m.* ***Right panel****: TMR single molecule data for EGF-TMR in vitro and in a live cell at the end of the photobleach, post EGF binding and taken from colocalized EGF-EGFR foci. Arrows indicate mean and s.e.m.* ***Inset****: live cell EGF-TMR photobleach steps after EGF has been added, taken from colocalized EGF-EGFR foci, with raw data traces (blue) and Chung-Kennedy-filtered traces (red). The difference between peak in vitro and in vivo foci intensity is consistent with in vivo GFP being marginally further (ca. 10-20 nm) from the coverslip surface in the TIRF evanescent field.*

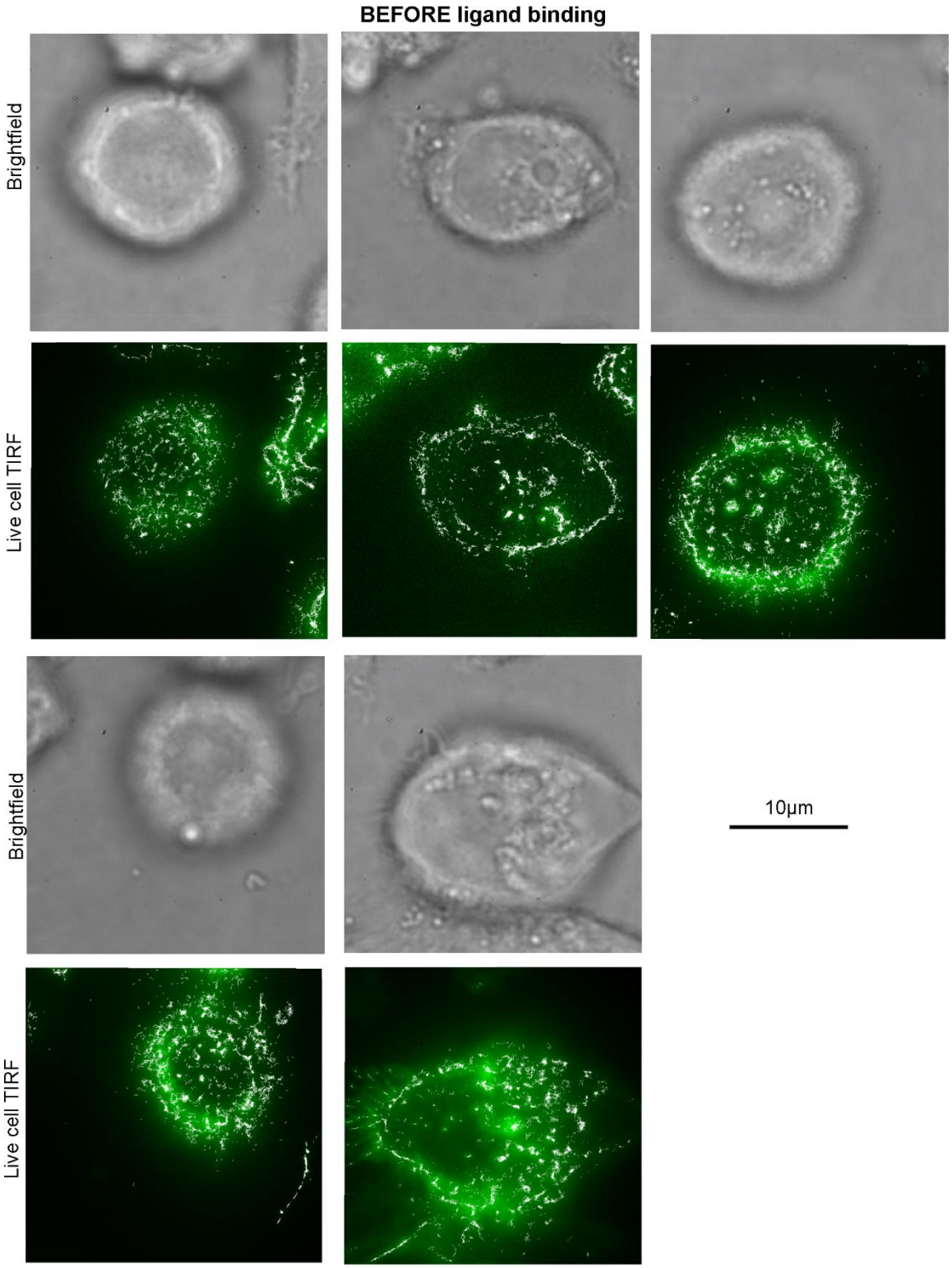


Figure S6: More examples of cells before addition of EGF ligand. *Brightfield images (grey) and TIRF images (green) shown with overlaid fluorescent foci tracking output (white).*

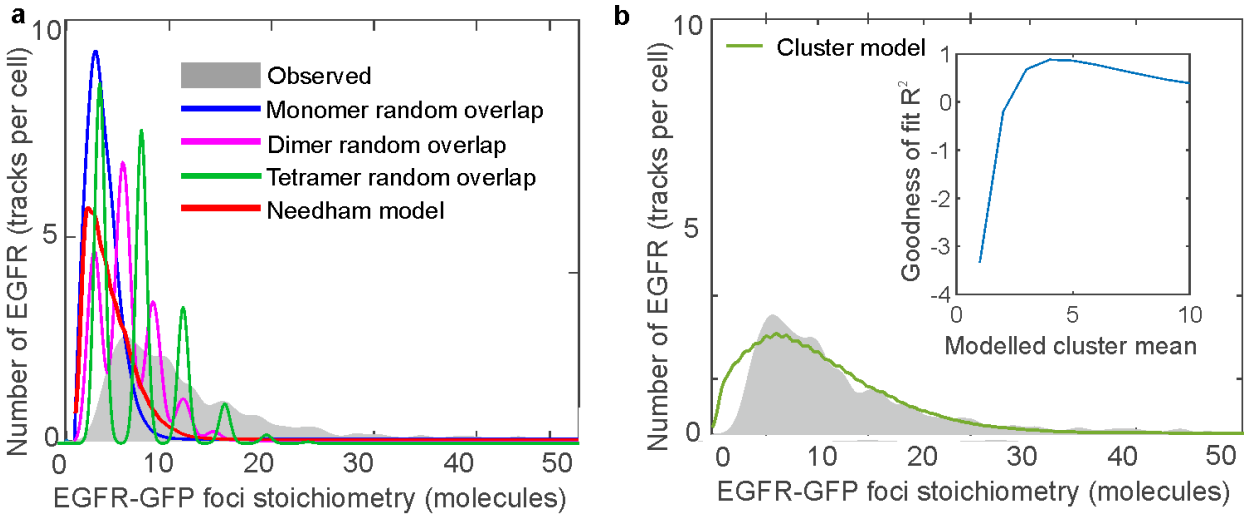


Figure S7 Random foci overlap model. a: *Model for the random overlap of EGFR molecules/cluster on the cell surface based on chance colocalization. The predicted cluster stoichiometry distributions are shown for the random overlap of EGFR monomers (blue), dimers (magenta), tetramer (green) and for a mixed model assuming 51.3% 1-mer, 21.3% 2-mer, 10.3% 3-mer, 5.3% 4-mer and 4.1% 5-mer per EGFR cluster) suggested by a previous single-molecule study (red) by Needham et al.* (39)*. Our own experimental stoichiometry distribution (grey) is in poor agreement with these models ($R^2<0$).* ***b:*** *Predictions (green) using a cluster model with expected average value of 4 molecules per EGFR cluster, together with experimental data (grey). A regression fit can account for ca. 90% of the experimental variance ($R^2=0.88$). R^2 as a function of average model cluster size is shown in the insert.*

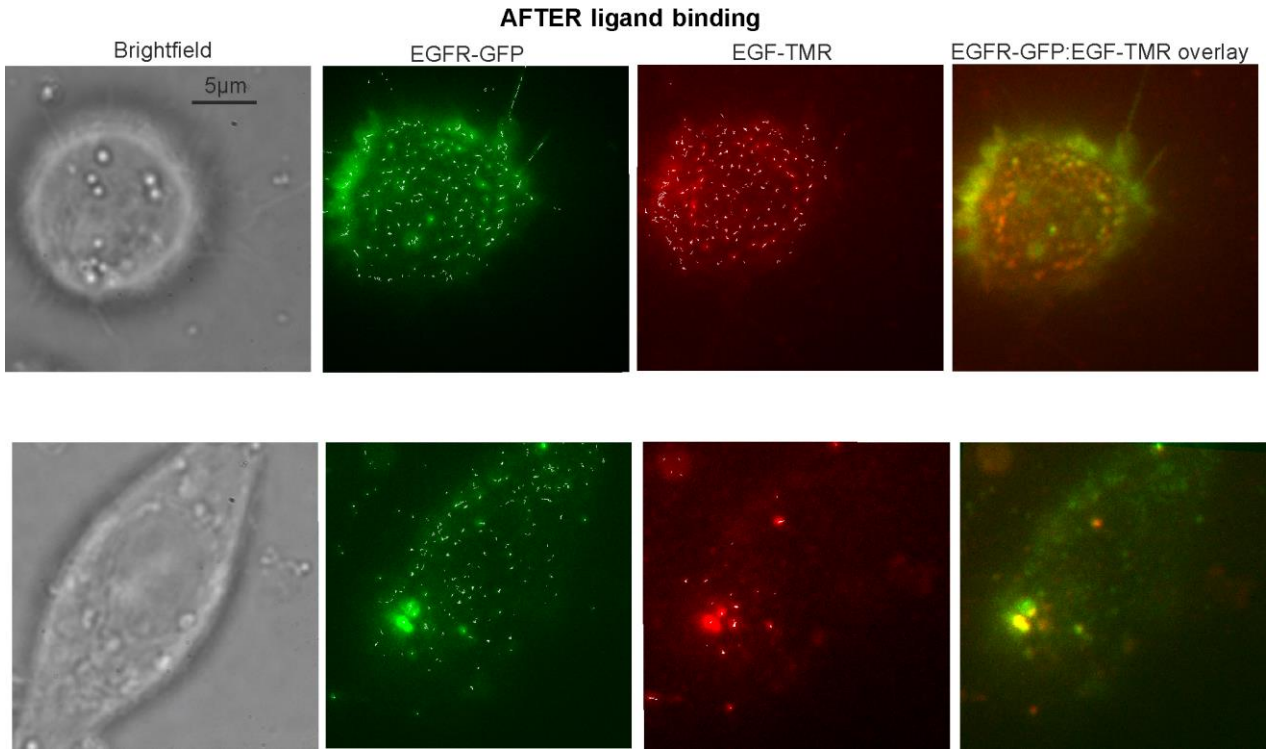


Figure S8: Colocalization of EGFR and EGF foci after addition of EGF. *Two examples (top and bottom) of cells taken ~10 min after addition of EGF. Shown from left to right are: brightfield images (grey), green channel with EGFR-GFP localizations (green), red channel with EGF-TMR localizations (red), and overlay of green and red channels, with yellow indicating regions of high colocalization.*

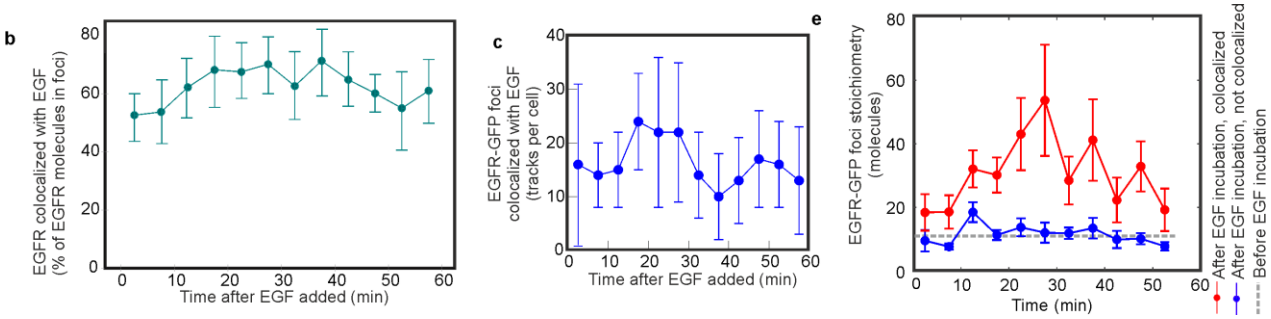
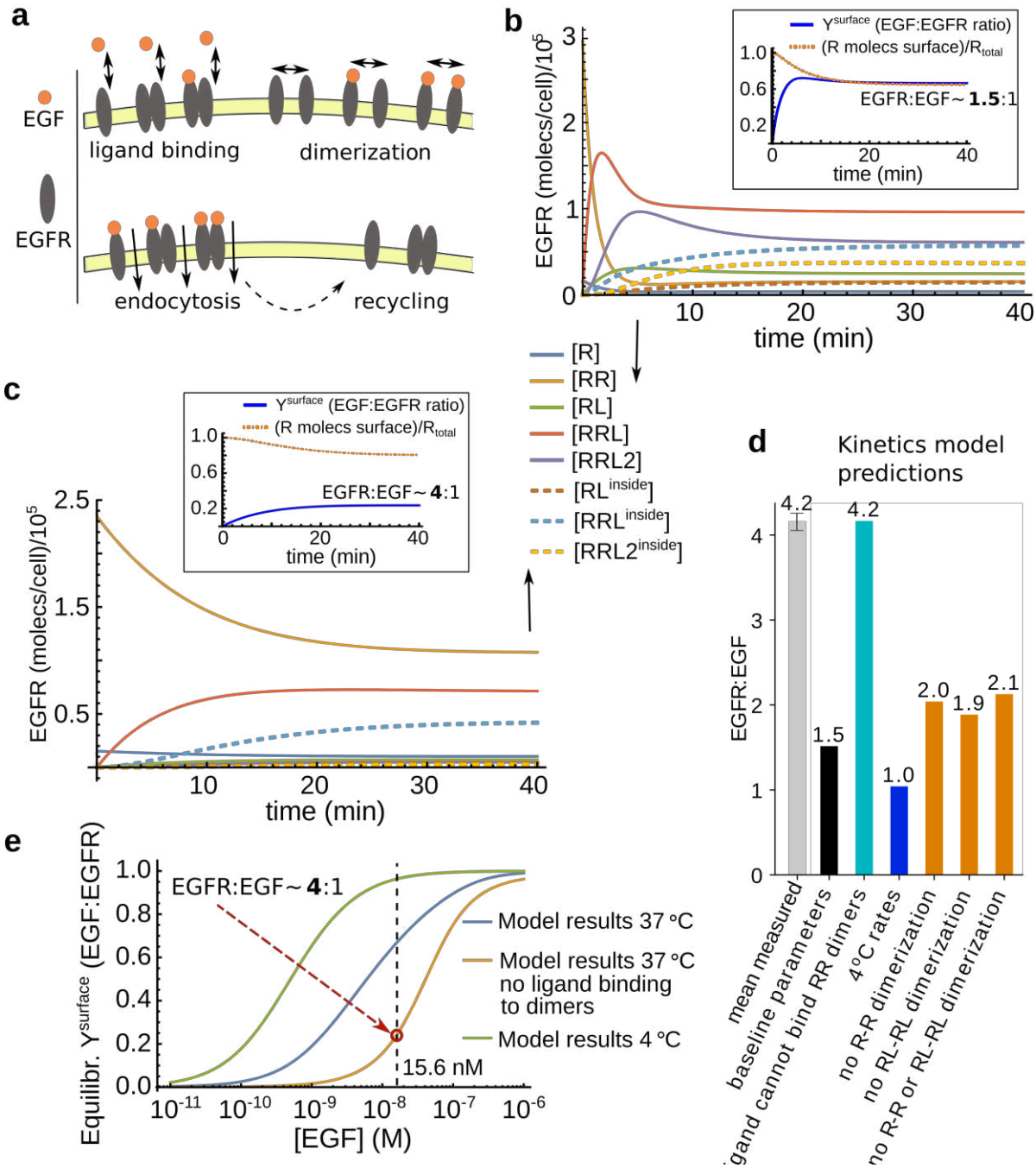
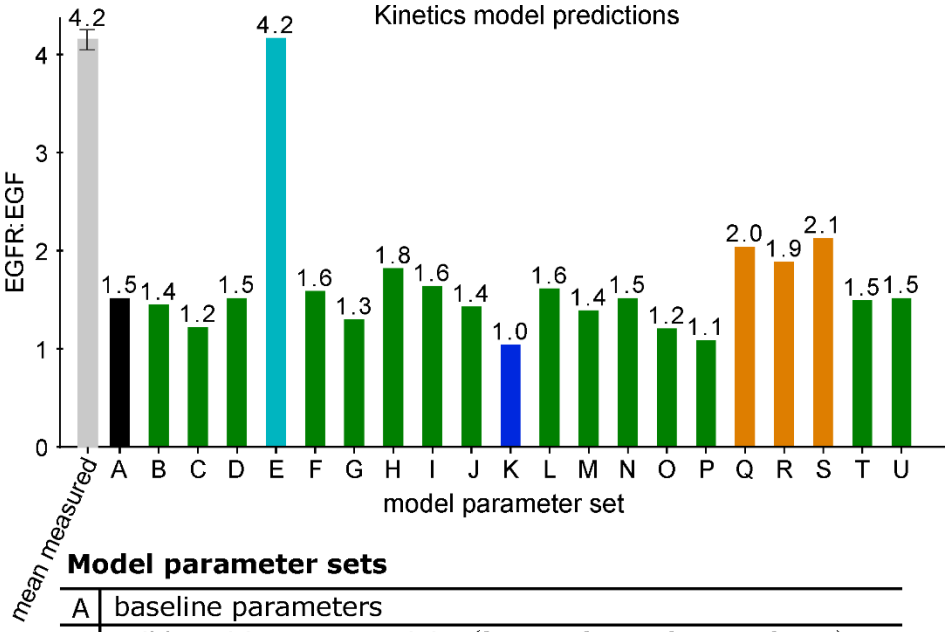


Figure S9: Characterization of EGFR and EGF foci stoichiometry after addition of EGF. (a) *Percentage of EGFR tracks colocalized to EGF.* ***(b)*** *Number of colocalized EGF-EGFR tracks detected per cell (with s.d. error bars) versus time* ***(c)*** *EGFR-GFP foci stoichiometry (number of molecules per foci) versus time (with s.d. error bars) for EGF-colocalized and not colocalized after EGF addition, and before EGF addition.*

Figure S10. Time-dependent EGFR-EGF kinetics model explains observations if EGF cannot bind receptor dimers. ***(a)*** *Schematic of reactions considered by the model: ligand binding to receptor monomers/dimers, receptor dimerization, endocytosis, and recycling.* ***(b)*** *Results of kinetics model for time dependence of receptor concentrations (for ligated and unligated monomers and dimers, and internalized components) choosing baseline parameters corresponding to 37°C (see Figure S11).* ***Inset:*** *fractional saturation $Y^{surface}$ equal to EGF:EGFR ratio (we experimentally measure its inverse, EGFR:EGF ratio).* ***(c)*** *Model predictions for same parameters as in (a) but assuming ligand binds only to monomers.* ***(d)*** *Model predictions for EGFR:EGF ratio (inverse of $Y^{surface}$) for several conditions (see Figure S10). Higher ratios in agreement with our data (grey bar) are obtained when ligand cannot bind to dimers (light blue).* ***(e)*** *Equilibrium $Y^{surface}$ versus EGF concentration for parameters shown in (a) and (b), as well as for reaction rates at 4°C). Black dashed line: experimental EGF concentration. Red dashed arrow: equivalent mean value of EGFR:EGF that we measure experimentally.*





Model parameter sets

A	baseline parameters
B	mild positive cooperativity ($k_{22,\mathrm{on}} = k_{11,\mathrm{on}},\ k_{22,\mathrm{off}} = k_{11,\mathrm{off}}$)
C	strong positive cooperativity ($k_{22,\mathrm{on}} = 10k_{11,\mathrm{on}},\ k_{22,\mathrm{o}k} = k_{11,\mathrm{o}k}$)
D	extreme negative cooperativity ($k_{22,\mathrm{on}} = 0$)
E	ligand cannot bind to RR dimers ($k_{21,\mathrm{on}} = 0$)
F	ligand cannot bind to monomers ($k_{11,\mathrm{on}} = 0$)
G	all ligand binding on-rate constants increase x2
H	all ligand binding on-rate constants decrease %2
I	all ligand binding off-rate constants increase x2
J	all ligand binding off-rate constants decrease %2
K	4°C rates
L	increased endocytosis rate (k_{endoc}=10%/min)
M	reduced endocytosis rate (k_{endoc}=3%/min)
N	increased recycling rate ($k_{recycle}$=40%/min)
O	zero recycling rate ($k_{recycle}$=0%/min)
P	no R-RL (hetero) dimerization ($l_{21,on}$=0)
Q	no R-R (homo) dimerization ($l_{20,on}$=0)
R	no RL-RL (homo) dimerization ($l_{22,on}$=0)
S	no R-R or RL-RL (homo) dimerization ($l_{20,on}$=0, $l_{20,on}$=0)
T	reduced total no. of receptors (R_{total}=200,000)
U	increased total no. of receptors (R_{total}=400,000)

Figure S11. Kinetics model predictions for various parameter sets. *Predictions of EGFR:EGF ratio (inverse of $Y^{surface}$) for baseline parameters (black bar, see details below) as well as for parameter sets (B to U) corresponding to a change with respect to the baseline parameters as indicated in the table below the graph. The plot shows that increased EGFR:EGF ratios closer to the experimentally measured value ~4 (grey bar) are predicted by the model for strongly reduced (absent) ligand binding to RR dimers (light blue bar). Reducing R-R and RL-RL homo-dimerization also leads to somewhat increased EGFR:EGF ratios ~2 (orange bars). Use of 4°C reaction rates leads to EGFR:EGF ~ 1 (dark blue bar). Baseline parameters are: R_{total}=310,000 molecules, [L]=15.6 nM; $k_{11,on} = 10^6 M^{-1}s^{-1}$, $k_{11,off} = 10^{-3}s^{-1}$, $k_{21,on} = k_{11,on}$, $k_{21,off} = k_{11,off}$, $k_{22,on} = k_{11,on}/7$ and $k_{22,off} = 2 \times k_{11,off}$ so that $K_{11} = 10^9 M^{-1}$, $K_{21} = K_{11}$, $K_{22} \approx K_{11}/14$; $l_{20,on} = 10^{-4}(molecules/cell)^{-1}s^{-1}$, $l_{20,off} = 10^{-1}s^{-1}$, $l_{21,on} = l_{20,on}$, $l_{21,off} = l_{20,off}$, $l_{22,on} = l_{20,on}/5$ and $l_{22,off} = 2 \times l_{20,off}$, so that $L_{20} = 10^{-3}(molecules/cell)^{-1}$, $L_{21} \approx L_{20}$ and $L_{22} \approx L_{20}/10$; $k_{endoc} = 6\,\%/min$, $k_{recycle} = 10\,\%/min$.*

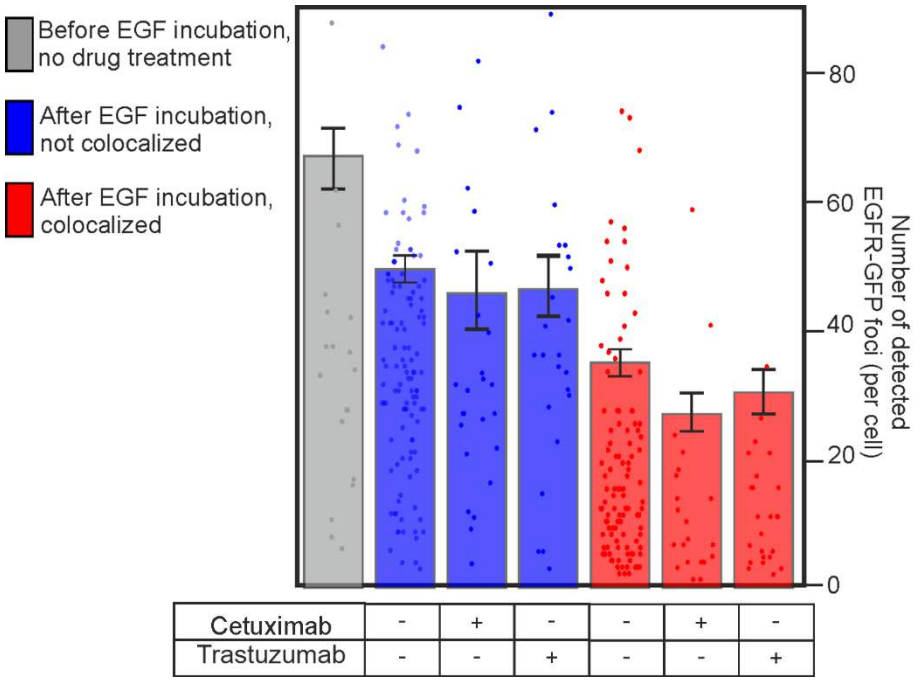


Figure S12: Number of EGFR-GFP foci detected per cell*. EGF-colocalized EGFR foci (red) and non-colocalized EGFR foci (blue) are shown. The table below indicates + or - for addition or not of anti-cancer drugs cetuximab and trastuzumab. Data before EGF incubation in the absence of drugs is shown as a grey bar. Error bars are s.d, N=10-117 cells per dataset.*

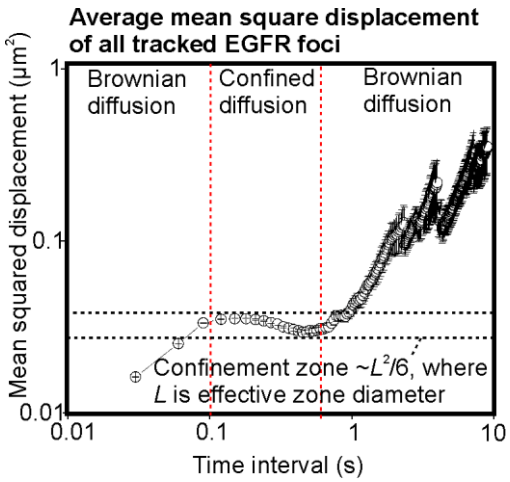


Figure S13: EGFR foci average mean square displacement vs time. *Log-log plot for average mean square displacement vs. time interval for all collated EGFR-GFP foci tracks before addition of EGF, putative confinement zone indicated (dashed lines), from number of foci N=770, acquired from number of cells N=19.*

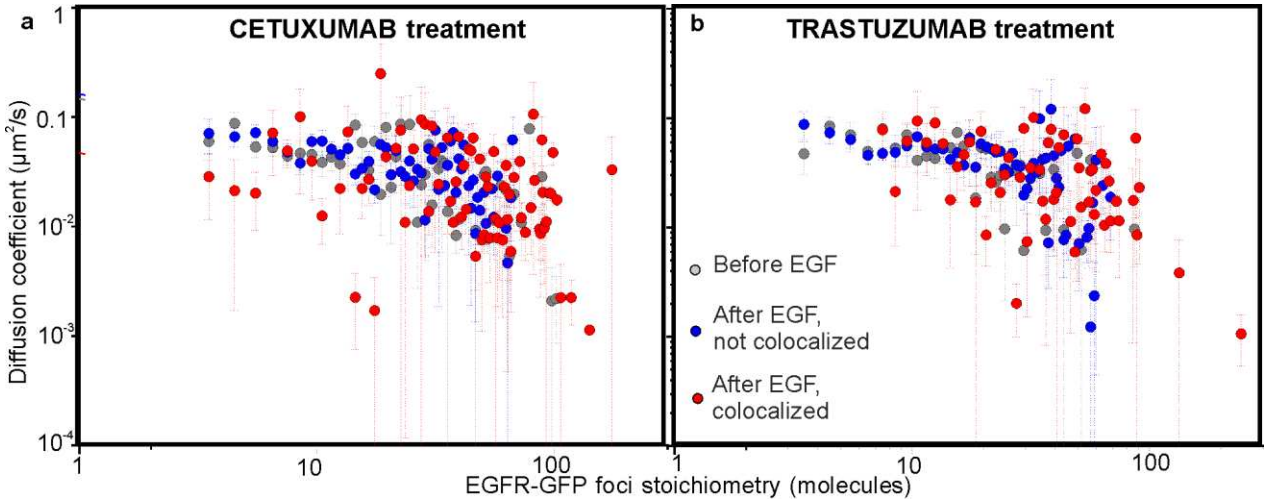


Figure S14: EGFR foci diffusion and anti-cancer drug treatment effects. *Log-log plots for variation of apparent diffusion coefficient D with EGFR stoichiometry S, for cells treated with drugs cetuximab* ***(a)*** *and trastuzumab* ***(b)****.*

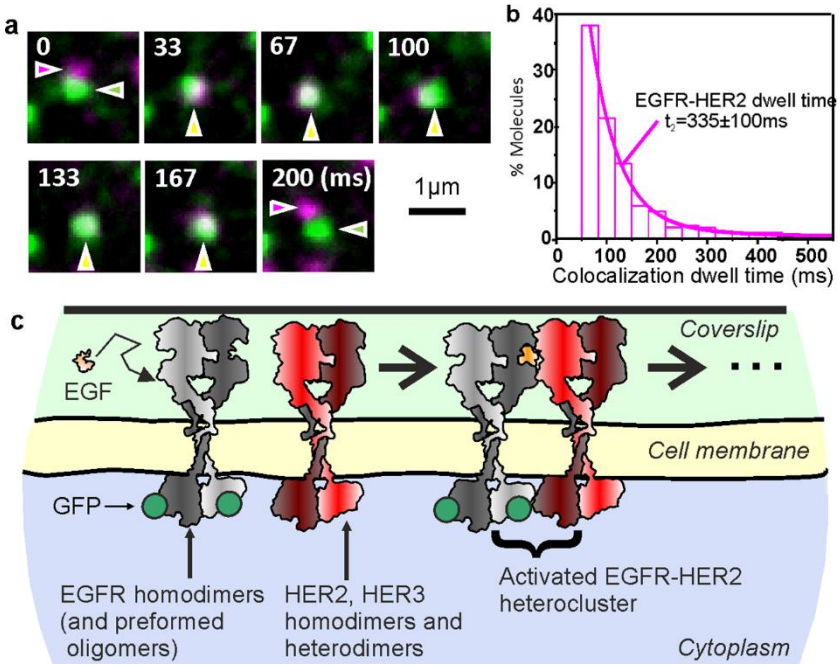


Figure S15. HER2-EGFR interactions*. **(a)** Single-molecule TIRF images of EGFR-HaloTag650 (magenta arrows) and HER2-GFP (green arrows) undergoing transient colocalization and co-diffusion (yellow arrows) in CHO-K1 cells. Time since start (ms) is indicated.* ***(b)*** *Histogram for the dwell time of colocalized EGFR/HER2 foci. A double exponential fit to these data is overlaid,*

and the mean and error (95% confidence bound) indicated for the average colocalization dwell time between HER2 and EGFR. Data extracted from N=4 cells corresponding to 400 colocalization detected events with random apparent colocalization dwell times assessed using 285 detected events. ***(c)*** *Schematic illustrating how HER2/HER3 and EGFR dimers might associate following EGF ligation, although we do not exclude monomer receptors also interacting.*

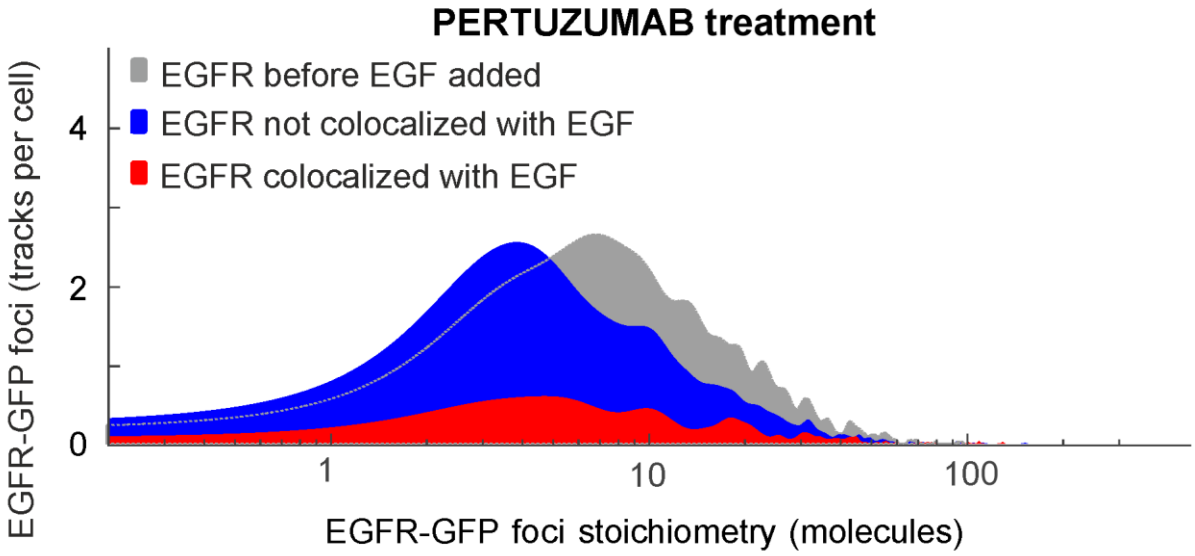


Figure S16. Effect of pertuzumab on EGFR foci stoichiometry. *Distribution of EGFR foci stoichiometry for cells treated with pertuzumab or, showing pre (grey) and post EGF addition for EGF-EGFR (red) and unligated EGFR (blue) foci, data collated across 60 min EGF incubation time. Number of cells per dataset in the range N =21.*

Supplementary Tables

Biochemical intervention				**EGFR foci stoichiometry, not colocalized with EGF**		**EGFR foci stoichiometry, colocalized with EGF**		N cells	N tracks/cell
E	C	T		Mean ± s.e.m (molecules per EGFR cluster)	N tracks	Mean ± s.e.m (molecules per EGFR cluster)	N tracks		
-	-	-		12.8±0.4	1,250	N/A	N/A	19	66
+	-	-		10.8±0.2	4,741	31.1±1.1	1,969	117	57
-	+	-		19.9±1.0	531	N/A	N/A	10	53
-	-	+		15.3±0.7	408	N/A	N/A	10	41
+	+	-		18.8±0.5	916	51.0±2.1	303	25	37
+	-	+		16.8±0.4	1,273	44.2±2.4	334	27	47

Table S1. Mean EGFR cluster stoichiometry. *Total number of tracks (N tracks) and number of individual cells (N cells) in the datasets are indicated. Biochemical interventions are: incubation with EGF (E), treatment with drug cetuximab (C) or treatment with drug trastuzumab (T), shown on the left-hand-side columns (- for no intervention, + for intervention).*

Supplementary movie legends

Movie S1. Single-colour TIRF imaging of live transfected SW620 cells. TIRF movie showing two adjacent cells transfected with EGFR-GFP (green) before addition of EGF.

Movie S2. Dual-colour TIRF imaging of live transfected

SW620 cells. TIRF movie showing a single cell transfected with EGFR-GFP (green), 10 minutes post addition of EGF-TMR (red, 100 ng/ml).

Movie S3. Dual-colour TIRF imaging of live CHO-K1 cell. TIRF movie showing a single CHO-K1 cell transfected with GFP-labelled HER2 (green) and HaloTag650-labelled EGFR (magenta).

Movie S4. Zoom into dual-colour TIRF imaging of live CHO-K1 cells. Zoom-in of cell shown in movie S3 displaying transient co-diffusion of HER2 and EGFR molecules.