Nanoparticle vesicle encoding for imaging and tracking cell populations

Paul Rees[1,2], John W Wills[1], M Rowan Brown[1], James Tonkin[1], Mark D Holton[1], Nicole Hondow[3], Andrew P Brown[3], Rik Brydson[3], Val Millar[4], Anne E Carpenter[2] & Huw D Summers[1]

For phenotypic behavior to be understood in the context of cell lineage and local environment, properties of individual cells must be measured relative to population-wide traits. However, the inability to accurately identify, track and measure thousands of single cells via high-throughput microscopy has impeded dynamic studies of cell populations. We demonstrate unique labeling of cells, driven by the heterogeneous random uptake of fluorescent nanoparticles of different emission colors. By sequentially exposing a cell population to different particles, we generated a large number of unique digital codes, which corresponded to the cell-specific number of nanoparticle-loaded vesicles and were visible within a given fluorescence channel. When three colors are used, the assay can self-generate over 17,000 individual codes identifiable using a typical fluorescence microscope. The color-codes provided immediate visualization of cell identity and allowed us to track human cells with a success rate of 78% across image frames separated by 8 h.

The ability to analyze collective behavior within a cell population[1] is crucial to the understanding of both normal and pathological human physiology. With the advent of high-throughput imaging technologies, the acquisition of live, single-cell images at subcellular resolution has enabled the study of statistically relevant populations in excess of 10^5 cells[2–4] over several days. For these studies, it is also essential to record quantitative information on individual cells, as cellular variability is a crucial aspect of population characteristics[5,6]. There are many situations in which uniquely identifying and tracking individual cells within a population is desirable. A protocol might require the identification of a rare cell within a population—for example, a stem cell or a cell within a side population[7]—or one may want to study the inherent heterogeneity of a population as a whole[8].

Colloidal quantum dots (QDs) are widely used as fluorescent markers for biological cells[9]. These particles are optically stable and biocompatible with the cellular environment[10,11]. The wavelength of the QD fluorescence can be tuned across the optical spectrum by varying the size, structure and chemical composition of the colloidal particles[12,13]. Initially, colloidal dots were used as direct replacements of existing fluorescence probes, but more recently, applications have emerged that exploit the unique properties of these nanoparticles[14,15]. The tunability of emission wavelengths for QDs has led to the suggestion that these particles could be used for optical barcoding, to provide a unique identifier for individual cells[16]. This has been achieved using analog coding techniques involving multicolor labeling and spectral discrimination; cell subtypes are identified by their spectral peak[17–19]. However, despite the flexibility in the choice of the peak wavelength, the emission spectrum is broad, thereby restricting the number of available codes and the utility of these approaches. Ultimately, all such barcoding approaches are also limited in that an exogenous label is used to identify preselected cell types or subgroups. For recognition of phenotypic variation that is a priori unknown to the experimenter, an innate label is required.

We present an approach using commercial QDs that exploits stochastic uptake dynamics of the nanoparticles in a simple assay that, with a standard laboratory fluorescence microscope, can individually label and track very large numbers of cells simultaneously. The uptake of nanoparticles and their subsequent encapsulation within vesicles inside the cell is a random process[20]; consequently, the number of nanoparticle-labeled vesicles (NLVs) varies from cell to cell. The stochastic processes driving nanoparticle endocytosis provide an endogenous cell-coding mechanism capable of labeling thousands of cells with distinct optical signatures. The endocytic pathway is critical to a host of cell functions such as metabolism, development, mobility and cell signaling[21]. Therefore, not only does this approach provide cell identification because the nanoparticle label is unique to each cell, but it is also informative of biological state and enables the study of the heterogeneity and biological consequences of endocytosis.

RESULTS

Nanoparticle-loaded vesicle encoding

We loaded commercially available QDs (CdSe-ZnS core-shell Qtracker 525, CdSe-ZnS Qtracker 585 and CdTe-ZnS Qtracker 705) sequentially into U-2 OS (human osteosarcoma) cells during 30-min exposure periods interspersed by particle wash-out and incubation periods of 30 min (Online Methods). The QDs are

[1]Centre for Nanohealth, School of Engineering, Swansea University, Swansea, UK. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Institute for Materials Research, School of Process, Environmental and Materials Engineering, University of Leeds, Leeds, UK. [4]General Electric Healthcare, The Maynard Centre, Cardiff, UK. Correspondence should be addressed to P.R. (p.rees@swansea.ac.uk).

RECEIVED 31 JANUARY; ACCEPTED 25 AUGUST; PUBLISHED ONLINE 14 SEPTEMBER 2014; DOI:10.1038/NMETH.3105

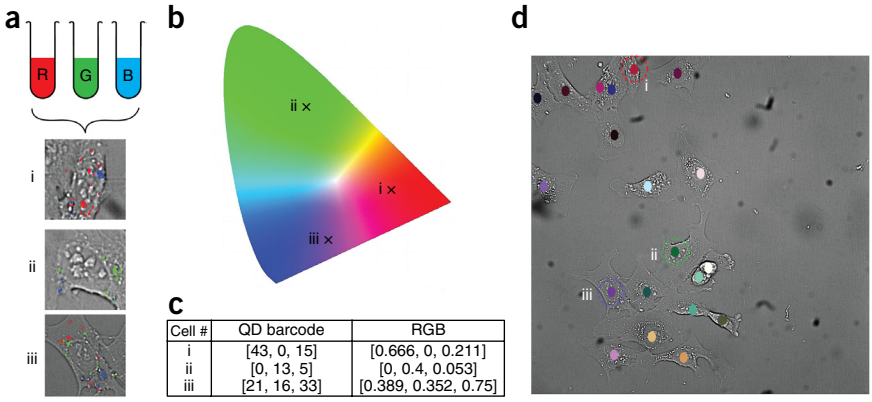

Figure 1 | Cell coding by using uptake of sequentially loaded colloidal QDs. (**a**) By using three different-color colloidal QDs, we obtain three integer values—corresponding to the number of vesicles labeled with each color for every individual cell—that can be used as identifiers. (**b**,**c**) As three colors are used, the number of loaded vesicles can be converted to a value between 0 and 1 (Online Methods), which defines a unique color on a standard RGB color space. (**d**) Assuming that the numbers of each colored vesicle remain constant during the cell cycle, this code can be used to track cells using time-lapse microscopy and imaging flow cytometry and furthermore be visualized using the unique color derived from the color space (outlines indicate cells i,ii and iii).

coated with peptides that target cell surface receptors, facilitating rapid internalization within 15 min (**Supplementary Fig. 1**). The combined concentration of QDs used for loading was within the range typically used for cell tracking without perturbing the cell cycle[22]. The staged exposure protocol produced vesicles loaded with 525-nm, 585-nm or 705-nm QDs (**Fig. 1a**). The number of color-coded NLVs per cell (on average, between 8 and 14 of each color per cell) can be used as a unique cell identifier. We use a reference notation: red (R, 705 nm), green (G, 585 nm) and blue (B, 525 nm) QDs; note that these are not the actual colors of fluorescence. We use this notation to generate a three-digit code based on the vesicle number of each cell (Online Methods), allowing placement within an RGB color map and display of cell identity through coloration (**Fig. 1b–d**). The 3-bit code was based on the use of three fluorescence channels; the remaining detection channels of the microscope are therefore available for additional fluorescent reporters.

The approach relies on minimizing the chance of a single vesicle containing QDs of more than one color. To confirm that all QDs of one color were internalized before exposure to the next color, we imaged cells using transmission electron microscopy (TEM). All vesicles contained only dots of the same emission wavelength, as identified by characteristic particle shape and contrast level in bright-field TEM (**Supplementary Fig. 2**). To further verify this, we confirmed the distinct chemical content of each dot type by energy-dispersive X-ray analysis (EDX; (**Supplementary Fig. 2**). We also tested uptake of nanoparticles in four additional cell lines, and they all supported labeling in this manner (**Supplementary Fig. 3**).

Code uniqueness

Across two independent experiments, we imaged 1,600 fields of view, identified an average of 39,087 cells and measured the number of loaded vesicles in each cell (Online Methods). The distributions of NLVs per cell for all three colors were similar, following a negative binomial, with mean values of 8–14 vesicles of each color per cell (**Fig. 2a–c**). We observed moderate correlations (Pearson coefficient of 0.35–0.47) between the numbers of NLVs of each color within each cell, i.e., the bit entries for the three-digit code (**Fig. 2d–f**); thus, the sequential labeling of the different QDs was not completely stochastic. A consequence of high correlation is that the conditional probability of finding a cell with a particular code combination increases. The observed correlation between NLVs of different colors serves to reduce the upper theoretical limit of RGB codes for the cell population compared with true Markovian random loading. Also, stochastic loading of the QDs dictates that cells randomly sample from this catalog of codes, with some combinations appearing more than once and some not at all; therefore, the maximum code number was not realized in practice.

Of the average 39,088 cells imaged per experiment, 17,015 cells had a unique barcode (43% of the measured population), 8,955 cells had one other cell within the population with the same code, 5,564 cells had two other cells within the population, and so on (**Fig. 2g**). However, spatial resolution from frame to frame enables cells that share a single code to be distinguished. From 1,600 fields of view, we selected a grid of 16 × 20 fields and mapped the centroid positions of each cell, approximately 7,760 cells in total (**Supplementary Fig. 4**.) In this example the chance of finding two cells with the same barcode in the same field of view was less than 2%, and the next duplicate cell would appear, on average, eight fields of view away (**Supplementary Fig. 4a**).

Improving the coding capacity: theoretical estimation

We have developed this assay with three colors because the resulting number of codes is sufficient for most applications and because it facilitates the convenient RGB color visualization. It is a trivial task to modify the loading protocol to include further colored QDs and expand the coding capacity. However, care must be taken not to introduce too high a QD dose, which may alter the behavior of the cells. When the loading dose per QD was reduced to 5 nM (**Supplementary Fig. 5**), using three-color labeling we obtained approximately 10% of the population with individual codes. To explore the possible number of codes that could be generated by introducing further colored QDs, we developed a simple algorithm (Online Methods) that generates random numbers from n negative binomial distributions, as used to fit the NLV histograms (**Fig. 2a–c**), plus a predefined correlation. We note that the total number of NLVs is kept constant by varying the mean at each color, and it is therefore the increased colors that increases the number of codes. We then simulated the loading of 39,088 cells with n different-colored QDs by generating integer random numbers from the negative binomial distributions. As expected, the number of distinct codes decreases as the correlation between the numbers of vesicles loaded with each color per cell increases, and our observed data point fell where predicted on the curve for a three-color exposure (**Fig. 2h**). The percentage of distinct codes is estimated to

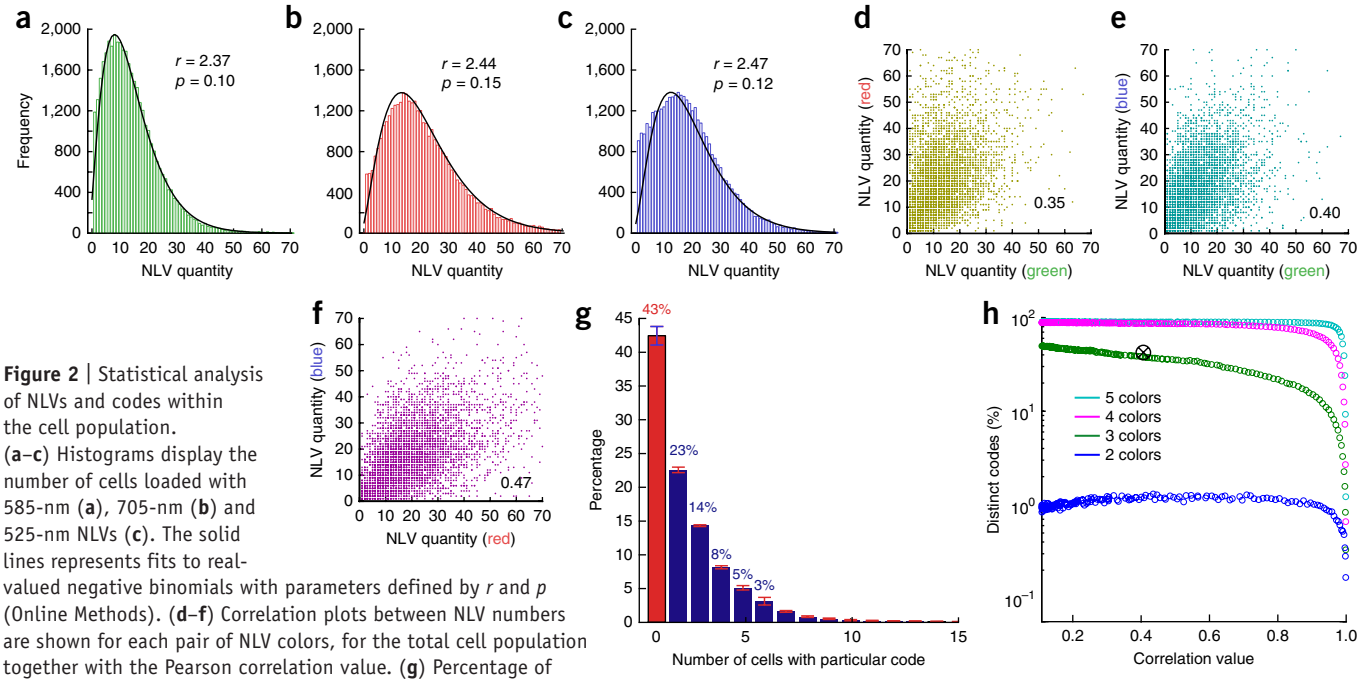

Figure 2 | Statistical analysis of NLVs and codes within the cell population. (**a**–**c**) Histograms display the number of cells loaded with 585-nm (**a**), 705-nm (**b**) and 525-nm NLVs (**c**). The solid lines represents fits to real-valued negative binomials with parameters defined by *r* and *p* (Online Methods). (**d**–**f**) Correlation plots between NLV numbers are shown for each pair of NLV colors, for the total cell population together with the Pearson correlation value. (**g**) Percentage of cells that share the same code with other cells in the population in $n = 2$ independent experiments. Error bars, s.d. (**h**) Theoretical estimate of the number of unique codes generated for different numbers of colors of QD load and different correlations between those loadings (Online Methods). The NLV distribution of the three-color QD loadings is calculated using a negative binomial distribution (Online Methods) with the parameters $r = 2.4$ and $p = 0.12$, which represents the average of the distribution values in **a**–**c**. The black × shows our actual experimental results in this paper (i.e., 17,015 unique codes observed at a correlation value of 0.407, which represents the mean correlation observed between the three colors in **d**–**f**).

dramatically increase from a maximum of 1.3% for two colors to 45% for three colors; for four or more colors, every one of the 39,088 cells is uniquely coded.

Visualizing heterogeneity in cell populations

We tested the ability of nanoparticle vesicle encoding to identify individual U-2 OS cells within large fields of view incorporating thousands of cells and thus help visualize heterogeneity in these populations. We mapped the heterogeneity in cellular endocytic activity (**Fig. 3a**) using the 3-bit NLV code to generate a false color for each cell from the RGB color map. A contour plot (**Fig. 3b**) illustrates the confluency of the cells. We measured the total number of reporting endosomes for sets of 160 cells in regions of high and low confluency. Histograms of particle number per cell clearly demonstrate that heterogeneity of QD uptake was correlated with confluency (**Fig. 3c**). We identified a subgroup of cells with anomalously high loading and therefore high values in the RGB code (particularly the 525-nm NLV) in regions of lowest cell confluency (**Fig. 3d**). This subgroup consisted of polyploid cells, which progress through the cell cycle but do not divide at mitosis; they are therefore abnormally large and have correspondingly large numbers of NLVs (**Supplementary Fig. 4**). This is to be expected as it has been shown for certain cell lines the uptake of particles is proportional to the cell area[23]. Thus, multicolor barcoding identifies cells of a distinct phenotype, locates and visually displays them within a large spatial field, and provides the correlating metric (in this case, NLV load to cell density) to highlight the population context–dependent cell heterogeneity. Elucidation of any causal relationship between cell density and ploidy would need further study.

Tracking cells in time-lapse imaging

Phototoxicity and photobleaching are major issues in live-cell time-lapse imaging. Frequent exposures are often necessary for successful tracking of cell identity but can seriously perturb cell physiology over time. We previously demonstrated that the pattern of QD-loaded vesicles of a single color can be used to track cells in time-lapse microscopy[24]; however, this still required image capture at short time intervals to track individual cells frame to frame. For this study, we used QD color-coding for extended time interval tracking in both U-2 OS cells observed with a high-throughput microscope (**Supplementary Fig. 6** and **Supplementary Video 1**) and A549 cells observed with a laser scanning confocal microscope (Online Methods). As time-lapse studies of cells usually require the measurement of fluorescent biomarkers, we also labeled the cell membrane with fluorescent wheat germ agglutinin (conjugated to Alexa 488), thereby demonstrating compatibility of the barcoding technique with other fluorescent markers (**Fig. 4a**).

Cell-tracking protocols rely on the identification of cells in successive time frames based on spatial correlations and trajectory[25]. The cell barcode provides an additional identifier for each individual cell that reduces the reliance on spatial position. This allows users to drastically increase the time between the capture of successive images and still maintain accurate identification and tracking of cells. We measured the cell NLV number for each color in an initial image frame (at time 0) and allocated a code to each cell. Then, after a chosen time interval, we measured the NLV for each cell from a second image frame and used a simple (root difference squared per color value, the L2-norm) metric between barcodes to relate individual cells. We compared the scoring success of the barcode technique relative to the ground

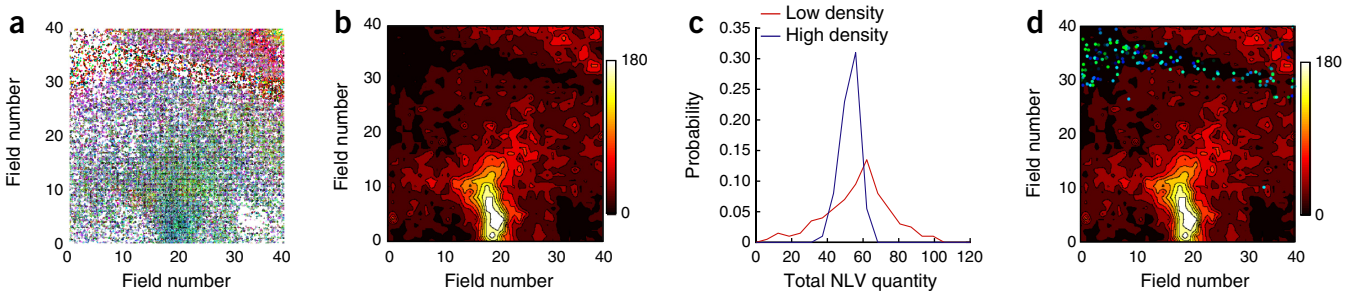

Figure 3 | Cell population barcoding for high-throughput microscopy. (**a**) Representation of a 40 × 40 array of microscopy fields of (13.3 mm × 13.3 mm), where each point represents one cell and the color denotes the RGB representation of the three-color NLV code. (**b**) Contour plot representing the density of cells (confluency); white represents high density and black represents low density for the fields shown in **a**. (**c**) Histograms of the number of coding endosomes per cell at high confluency (more than 60 cells per field of view; blue line) and low cell confluency (fewer than ten cells per field of view; red line). (**d**) Contour plot in **b** with overlaid data from a subpopulation of cells with enhanced particle uptake. Blue and green circles indicate the RGB color representation and cell centroid position.

truth obtained by manually tracking the cells using the corresponding bright-field images (**Fig. 4b**). Over 90% of the approximately 80–90 cells in the chosen image (**Supplementary Fig. 7**) could be tracked with a 4-h time-frame interval; and even with an 8-h gap between time frames, nearly 80% of cell were still identifiable (**Fig. 4b**). We compared our tracking method against the cell-tracking algorithms based on cell location and trajectory, which are freely available in CellProfiler[26]. Of the three available methods, the 'distance' (nearest neighbor) algorithm performed the best (**Supplementary Fig. 8**), and we compared this to our approach (**Fig. 4b**). Because all cells within a single image frame are tracked, our technique can be scaled to an arbitrary area size using multiple fields of view, as is standard with high-throughput microscopy. Notably, we use a matching algorithm only to identify the cell within the field of view that has the number of colored vesicles most closely matching that of the starting cell. Further improvement in accuracy and exposure intervals could be realized with more sophisticated algorithms or with optimization of the microscope.

Notably, the QDs appeared to persist in late endosomes or lysosomes throughout each cell's lifetime; when the cell undergoes mitosis the loaded vesicles are partitioned between the daughter cells. This partitioning, like the original loading, is random[27]; however, the overall vesicle number is conserved, and so the barcode identification can be continued across generations. As nondividing cells retain the same code, it is relatively simple to identify (within the remaining dividing fraction) which daughter cells originated from which parent cell (**Fig. 4c** and **Supplementary Fig. 6**). Daughter cells produced by the two mitotic events can be linked to their parent cells through a simple algorithm that bases the choice of parent-daughter on minimization of the summed difference between labeled endosomes in all three color channels (**Supplementary Fig. 6**).

For live-cell tracking the requirement is to identify cells within the bounded area of the image field. In this case, the stability of the code becomes the most important criterion, and it is also important to maximize the number of unique codes generated to ensure that neighboring cells are as distinct as possible. This ensures that any fluctuation in the measured barcode of the cells does not lead to confusion with neighboring cells. In our experiment we observed an s.d. in the fluctuation of each color-code of 1.5 NLVs between successive 1-h interval time frames (**Supplementary Fig. 9**), which consequently increased the probability of erroneously finding an identically labeled cell in the field of view (**Supplementary Fig. 10**). The fluctuation in the measured barcode observed in our experiment means that the next apparent duplicate cell appeared on average 1.3 fields of view away (compared with eight fields of view away when no error in measurement is present). However, as we compared cell barcodes within a single field of view, this did not prevent high-accuracy

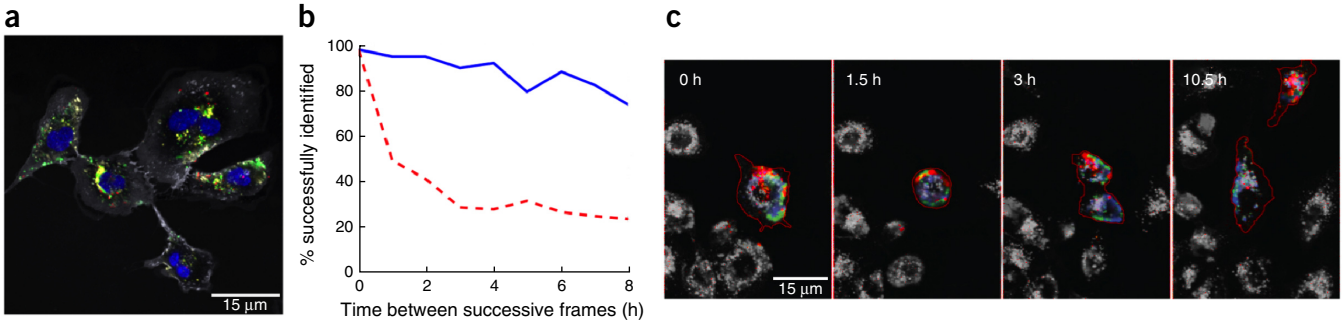

Figure 4 | Cell tracking over many hours by microscopy. (**a**) Fixed A549 cells loaded with 585-nm (yellow), 655-nm (green) and 705-nm (red) Qtracker QDs; wheat germ agglutinin conjugated to Alexa Fluor 488 (gray) to mark the cell membrane; and a Hoechst nuclear stain (blue). (**b**) Live cells labeled as in **a** were imaged using a laser scanning confocal microscope. The plot shows tracking of a single field of view with approximately 80–90 cells with our approach (blue) and the distance (nearest neighbor) method in CellProfiler (red) over the indicated time intervals. The success rates are relative to the ground truth obtained by manually tracking the cells using the bright-field images. (**c**) Progression of a QD-labeled cell undergoing a mitotic event (outlined in red).

tracking (**Fig. 4b**). Tracking was also possible when the number of QD colors was reduced from the three used in our experiment to a protocol labeling with two or even just one QD (**Supplementary Fig. 11**). There was a decrease in the accuracy of tracking over long time intervals with fewer QDs, consistent with our previous work, but this reiterates the applicability of the technique to instruments with limited fluorescence channels.

DISCUSSION

Using a separate loading of three different-colored QDs into cells in a population, we have developed a method that enables cells to self-label with unique identifiers, specified by the number of vesicles of each color within that cell. The assay requires no specialty nanoparticles or imaging instrumentation and can be interpreted using standard RGB visualization tools. It can be used accurately identify and track cells in time-lapse microscopy experiments, even with very low frame rates that minimize photobleaching and phototoxicity. Furthermore, the number of QDs taken up by a cell can also reveal information about cell function such as metabolism, cell cycle or the phenotype of individual cells within a population.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council, UK, under grant EP/H008683/1, 'Nanoparticle Cytometrics'. P.R. was supported by the Engineering and Physical Sciences Research Council, UK International Collaboration Sabbatical scheme under grant EP/J00619X/1 and the European Regional Development Funded, Swansea Centre for Nanohealth (CNH).

AUTHOR CONTRIBUTIONS

P.R., H.D.S., A.E.C. and M.R.B. designed the experiments; P.R., M.R.B., J.T. and V.M. analyzed the data; and P.R., A.E.C. and H.D.S. wrote the manuscript in close collaboration with the other coauthors. M.D.H. implemented the cell culture dosing regimen; J.W.W., M.R.B. and V.M. performed the microscopy measurements and the image analysis. N.H., A.B. and R.B. carried out the TEM and EDX studies. All coauthors discussed the results and approved the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Axelrod, R., Axelrod, D.E. & Pienta, K.J. Evolution of cooperation among tumor cells. *Proc. Natl. Acad. Sci. USA* **103**, 13474–13479 (2006).
2. Eliceiri, K.W. *et al.* Biological imaging software tools. *Nat. Methods* **9**, 697–710 (2012).
3. Perlman, Z.E. *et al.* Multidimensional drug profiling by automated microscopy. *Science* **306**, 1194–1198 (2004).
4. Rajaram, S., Pavie, B., Wu, L.F. & Altschuler, S.J. PhenoRipper: software for rapidly profiling microscopy images. *Nat. Methods* **9**, 635–637 (2012).
5. Snijder, B. *et al.* Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**, 520–523 (2009).
6. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125 (2011).
7. Ho, M.M., Ng, A.V., Lam, S. & Hung, J.Y. Side population in human lung cancer cell lines and tumors is enriched with stem-like cancer cells. *Cancer Res.* **67**, 4827–4833 (2007).
8. Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862 (2009).
9. Bruchez, M. Jr., Moronne, M., Gin, P., Weiss, S. & Alivisatos, A.P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
10. Jaiswal, J.K., Mattoussi, H., Mauro, J.M. & Simon, S.M. Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nat. Biotechnol.* **21**, 47–51 (2003).
11. Brown, M.R. *et al.* Long-term time series analysis of quantum dot encoded cells by deconvolution of the autofluorescence signal. *Cytometry A* **77**, 925–932 (2010).
12. Parak, W.J., Pellegrino, T. & Plank, C. Labelling of cells with quantum dots. *Nanotechnology* **16**, R9–R25 (2005).
13. Koo, O.M., Rubinstein, I. & Onyuksel, H. Role of nanotechnology in targeted drug delivery and imaging: a concise review. *Nanomedicine* **1**, 193–212 (2005).
14. Ferrari, M. Cancer nanotechnology: opportunities and challenges. *Nat. Rev. Cancer* **5**, 161–171 (2005).
15. Chan, W.C.W. *et al.* Luminescent quantum dots for multiplexed biological detection and imaging. *Curr. Opin. Biotechnol.* **13**, 40–46 (2002).
16. Alivisatos, P. The use of nanocrystals in biological detection. *Nat. Biotechnol.* **22**, 47–52 (2004).
17. Fournier-Bidoz, S. *et al.* Facile and rapid one-step mass preparation of quantum-dot barcodes. *Angew. Chem. Int. Ed. Engl.* **47**, 5577–5581 (2008).
18. Stsiapura, V. *et al.* Functionalized nanocrystal-tagged fluorescent polymer beads: synthesis, physicochemical characterization, and immunolabeling application. *Anal. Biochem.* **334**, 257–265 (2004).
19. Han, M., Gao, X., Su, J.Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat. Biotechnol.* **19**, 631–635 (2001).
20. Summers, H.D. *et al.* Statistical analysis of nanoparticle dosing in a dynamic cellular system. *Nat. Nanotechnol.* **6**, 170–174 (2011).
21. Scita, G. & Di Fiore, P.P. The endocytic matrix. *Nature* **463**, 464–473 (2010).
22. Errington, R.J. *et al.* Single cell nanoparticle tracking to model cell cycle dynamics and compartmental inheritance. *Cell Cycle* **9**, 121–130 (2010).
23. Wilhelm, C., Gazeau, F., Roger, J., Pons, J.N. & Bacri, J.-C. Interaction of anionic superparamagnetic nanoparticles with cells: kinetic analyses of membrane adsorption and subsequent internalization. *Langmuir* **18**, 8148–8155 (2002).
24. Tonkin, J.A. *et al.* Automated cell identification and tracking using nanoparticle moving-light-displays. *PLoS ONE* **7**, e40835 (2012).
25. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).
26. Carpenter, A.E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
27. Kim, J.A., Åberg, C., Salvati, A. & Dawson, K.A. Role of cell cycle on the cellular uptake and dilution of nanoparticles in a cell population. *Nat. Nanotechnol.* **7**, 62–68 (2012).

ONLINE METHODS

Quantum-dot loading and preparation of cells for high-throughput microscopy. For all experiments, U-2 OS (ATCC HTB-96) cells were maintained in McCoy's 5a medium supplemented with 10% fetal calf serum (FCS), 1 mM glutamine and antibiotics, and incubated at 37 °C in an atmosphere of 5% CO_2 in air. Cultures were examined using a Axiovert 25 light microscope (Carl Zeiss) and routinely tested for *Mycoplasma* using MycoAlert mycoplasma detection kit (Lonza). Cells were sequentially loaded with commercially available targeted nanocrystals using the Qtracker585 Cell Labeling Kit (Invitrogen, Q25011MP), Qtracker 705 Cell Labeling Kit (Invitrogen, Q25061MP) and Qtracker 525 Cell Labeling Kit (Invitrogen, Q25041MP) at a range of concentration as indicated in the main text. For all three different nanocrystal loadings, the Qtracker reagents A and B were premixed and then incubated for 5 min at room temperature. Fresh full growth medium (2 ml) was added to the tube and vortexed for 30 s. This labeling solution was then added to the cells in the order given in the text and then incubated for 30 min at 37 °C. After the 30-min period, the cells were washed twice with fresh medium to remove any free QDs from the cell medium. This process was repeated for the two other loadings.

High-throughput microscopy. The In Cell Analyzer 2000 (GE Healthcare) was used to measure the fluorescence signals from the cellular nucleus labeled via the DNA stain Hoechst 33342 (using the prescribed protocol: diluted 1:100 in sterile H_2O, 2 µl is added to 2 ml of serum-free medium to create a stain culture medium that is added to the cells and incubated for 30 min at 37 °C) and the three different QDs as described above. The QD fluorescence was acquired using 100-ms exposure times and the fluorescence filter sets QD525, FITC; QD585, Cy3; QD705, Cy5. The fluorescence properties of 39,088 cells were collected over 1,600 (40 × 40) fields of view using a 40× objective lens.

A watershed segmentation algorithm was used to segment individual nuclei on the basis of their fluorescent Hoechst 33342 signal and provide a binary mask. The cytoplasm mask of each cell was estimated by simultaneously dilating with a disk-shaped structural element of radius 30 µm. Dilation occurs only when neighboring cytoplasm masks are disjoint. Each cell mask is then applied to the fluorescence channels of each QD type, and the number of distinct NLVs within each mask is counted. We compared the NLV count results to those obtaining using a *k*-means cluster algorithm in Matlab (Matlab 2010, MathWorks) that gave identical results for NLV counts.

Counting NLV numbers per cell. To count the number of NLVs, we used a simple peak-find algorithm in Matlab that counts connected pixels above a threshold value. We chose a threshold value of 20% lower than the highest intensity value that gave highly reproducible values, which were fairly insensitive to this threshold value (**Supplementary Software**).

Generating RGB color-codes from NLV number. The RGB code is represented by three numbers between 0 and 1 for each of the red, green and blue color channels. We could simply linearly rescale the number of NLVs for each of the three colors from between 0 and the maximum number observed to a value between 0 and 1. However, because the NLV distribution shows cells with similar numbers of certain NLVs, better color differentiation is obtained if the cumulative distribution of the histograms in **Figure 2a–c** is scaled to give a value between 0 and 1.

Tracking cells using the NLV number for each channel. In order to track cells between successive time frames, we compare the number of NLVs for each color in every cell between the two time-frame images to obtain the closest match. We calculate the root difference squared per color value (the L2-norm) for each of the cells in the two images and the minimize this to allocate the closest match, which is designated as the same cell in both images (**Supplementary Software**).

Quantum-dot loading and preparation of cells for laser scanning confocal microscopy. Lung carcinoma epithelial A549 (ATCC CCL-185) cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum, 1% penicillin/streptomycin and 1 mM glutamine. Cultures were examined using an Axiovert 25 light microscope (Carl Zeiss) and routinely tested for *Mycoplasma* using MycoAlert mycoplasma detection kit (Lonza). Cells were seeded at 6,000 cells/cm^2 and then returned to the incubator for 12 h under standard conditions (37 °C, 5% CO_2, 95% humidity). Cells were then loaded with commercially available targeted nanocrystals at 10 nM concentration in the order Qtracker 585, Qtracker 655, Qtracker 705 (Life Technologies Q25011MP, Q25021MP, Q25061MP, respectively). This was conducted by mixing equal volumes of Qtracker reagents A and B, incubating for 5 min at room temperature then adding 2 ml complete DMEM and vortexing for 30 s. Cells were then sequentially incubated with the labeling solutions, consisting of 1 h exposure before washing out with three changes of complete DMEM and a further 1-h incubator recovery. This process was then repeated for the other two loadings. At the end of the final incubation, cells were washed thoroughly with prewarmed Hank's buffered salt solution (HBSS). A 7.5 mg/ml solution of wheat germ agglutinin–Alexa Fluor 488 conjugate (WGA) (Life Technologies W11261) in HBSS was then added for 10 min to label cell membranes. WGA is a live-cell marker that binds to the cell membrane and is used in a wide range bioassays including study of nuclear protein transport, mapping the distribution of gap junctions and determination of intracellular protein distributions. Cells were then washed and transferred into the microscope's live-cell incubator at standard conditions in complete phenol red–free DMEM for imaging.

Live-cell laser scanning confocal microscopy. 8-h time-lapse microscopy was performed at 30-min intervals under a 20× plan apochromatic objective yielding a 0.4-mm^2 field at 16-bit, 1,024 × 1,024–pixel resolution using an LSM 710 laser scanning confocal microscope (LSCM) (Carl Zeiss). Cells were maintained under standard conditions throughout. The four channels (WGA + three QDs) were collected sequentially to maximize signal and minimize the intensity of laser exposure to the cells. Image collection was carried out automatically using ZEN software version 2009 (Carl Zeiss). Singly labeled cells were subsequently used to confirm no fluorescence cross-talk between tracks under the image collection settings. When a membrane stain was used, we were able to segment robustly using a binary mask based on the membrane stain intensity.

doi:10.1038/nmeth.3105

Probability distributions. The NLV histograms were characterized by fitting to a negative binomial distribution given by

$$f(n;r,p)=\binom{n+r+1}{n}(1-p)^r\,p^n$$

where n is the number of loaded vesicles and the parameters r and p describe the shape of the distribution, i.e., the overdispersion. A negative binomial is chosen as it captures the overdispersion of NLV loading, which is observed in experiment and provides a theoretical model of coding capacity. Overdispersion represents an increased variability of the uptake compared with the expected random uptake process.

Generation of *N* sets of random numbers from an arbitrary probability distribution with a predefined correlation. To the best of our knowledge, there appears to be no closed-form algorithm to generate random numbers from N arbitrary distributions with a predefined correlation. We have used an approximate solution that takes the random numbers generated from a single distribution and modifies each random number generated by multiplying by a factor of λM, where λ is itself a random number between 0 and 1 and the parameter M then represents how correlated the distributions will be, i.e., the less correlated, the greater M becomes (a renormalization is used to ensure that the mean of the distribution remains constant). A simple evolutionary algorithm that varies M to obtain the desired correlation is then used. This method preserves the probability distribution required for the purposes of this paper. The algorithm used is outlined in the supplementary information together with the Matlab code used (**Supplementary Software**).