

Quantitative self-assembly prediction yields targeted nanomedicines

Yosi Shamay^{1,2}, Janki Shah¹, Mehtap Işık^{1,3}, Aviram Mizrahi^{1,4}, Josef Leibold¹, Darjus F. Tschaharganeh⁵, Daniel Roxbury⁶, Januka Budhathoki-Uprety¹, Karla Nawaly¹, James L. Sugarman¹, Emily Baut^{1,7}, Michelle R. Neiman¹, Megan Dacek^{1,7}, Kripa S. Ganesh^{1,7}, Darren C. Johnson^{1,3}, Ramya Sridharan^{1,7}, Karen L. Chu^{1,7}, Vinagolu K. Rajasekhar¹, Scott W. Lowe¹, John D. Chodera¹ and Daniel A. Heller^{1,7*}

Development of targeted nanoparticle drug carriers often requires complex synthetic schemes involving both supramolecular self-assembly and chemical modification. These processes are generally difficult to predict, execute, and control. We describe herein a targeted drug delivery system that is accurately and quantitatively predicted to self-assemble into nanoparticles based on the molecular structures of precursor molecules, which are the drugs themselves. The drugs assemble with the aid of sulfated indocyanines into particles with ultrahigh drug loadings of up to 90%. We devised quantitative structure-nanoparticle assembly prediction (QSNAP) models to identify and validate electrotopological molecular descriptors as highly predictive indicators of nano-assembly and nanoparticle size. The resulting nanoparticles selectively targeted kinase inhibitors to caveolin-1-expressing human colon cancer and autochthonous liver cancer models to yield striking therapeutic effects while avoiding pERK inhibition in healthy skin. This finding enables the computational design of nanomedicines based on quantitative models for drug payload selection.

A significant limitation of targeted nanoparticle drug carrier design^{1–3} is that complex synthetic schemes are often required, resulting in low loadings and higher barriers to clinical translation^{4–6}. Until recently, the process of drug encapsulation into stable nano-formulations has been largely experimentally based^{7,8}. However, recent reports on the prediction of nano-formulation *in silico* suggest that this goal will be a major contribution to the emergence of ‘nanoinformatics’, a new sub-field within nanotechnology⁹. In drug carrier design, quantitative structure–property relationship (QSPR) calculations have been used to predict colloidal drug aggregation^{10–12}, drug loading in lipid formulations¹³, and *in vivo* performance^{14–16}. Molecular dynamics simulations have been used to investigate nanoparticle supramolecular interactions^{17–19} and vehicle selection²⁰. However, quantitative approaches have not yet provided appreciable predictive power to enable the design of delivery vehicles *de novo* based on drug payload selection.

Herein, we found that a subset of sulfated indocyanine dyes self-assemble with hydrophobic drugs to form stable tumour-targeted nanoparticles wherein the formation can be accurately predicted using quantitative information from the structure of the encapsulated drug. Via quantitative structure-nanoparticle assembly prediction (QSNAP) calculations, we identified two molecular descriptors to predict which drugs would assemble with indocyanine into nanoparticles and nanoparticle size with an average accuracy of up to 15 nm. Moreover, this approach also revealed important molecular structure features that enable self-assembly and nanoparticle formation. Importantly, the resulting indocyanine nanoparticles (INPs) were found to encapsulate drugs with high loadings of up

to 90% by mass. We assessed the targeted drug delivery properties of two such nanoparticles, encapsulating the tyrosine kinase inhibitors sorafenib and trametinib. Strikingly, we found selective caveolin-1-mediated tumour uptake and exceptional net anti-tumour efficacies in a genetically modified mouse model for hepatocellular carcinoma and a xenograft model for human colorectal cancer. The nanoparticles prevented the inhibition of ERK phosphorylation in the skin, demonstrating that this targeting strategy exhibits strong therapeutic benefits and may obviate skin rash—a dose-limiting side-effect of kinase inhibitors.

Indocyanine-stabilized drug nanoparticles

We explored the parameter space of excipients that could facilitate self-assembly of drugs into colloidal-stable nanoparticles with high drug loadings. We estimated that a clinically-useful nanoformulation would contain at least 2 mg ml⁻¹ of drug suspended in water or PBS²¹. We designed a simple nanoprecipitation process to test the propensity of excipient molecules to stabilize hydrophobic drugs within sub-100-nm nanoparticles (Methods). We examined detergents, polyelectrolytes, and lipids previously used to form drug nanocrystals^{22,23}, as well as azo dyes, shown recently to form colloidal drug aggregates²⁴. We also tested water-soluble dyes of different classes, including phthalocyanines, triphenylmethanes and cyanines. We then chose an initial set of nine hydrophobic drugs (Supplementary Fig. 1a) with poor water solubilities (<10 µg ml⁻¹). We found that a subset of drugs was able to form stable nanoparticles with sulfated organic dyes from three different dye families: azo, triphenylmethane and indocyanine^{25,26} (Fig. 1a,b, Supplementary

¹Memorial Sloan Kettering Cancer Center, New York, NY, USA. ²Faculty of Biomedical Engineering, Technion—Israel Institute of Technology, Haifa, Israel.

³Tri-Institutional PhD Program in Chemical Biology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ⁴Department of Otolaryngology Head and Neck Surgery, Rabin Medical Center and Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. ⁵Helmholtz-University Group ‘Cell Plasticity and Epigenetic Remodeling’, German Cancer Research Center (DKFZ) & Institute of Pathology University Hospital, Heidelberg, Germany. ⁶Department of Chemical Engineering, University of Rhode Island, Kingston, RI 02881, USA. ⁷Weill Cornell Medical College, Cornell University, New York, NY, USA.

*e-mail: hellerd@mskcc.org

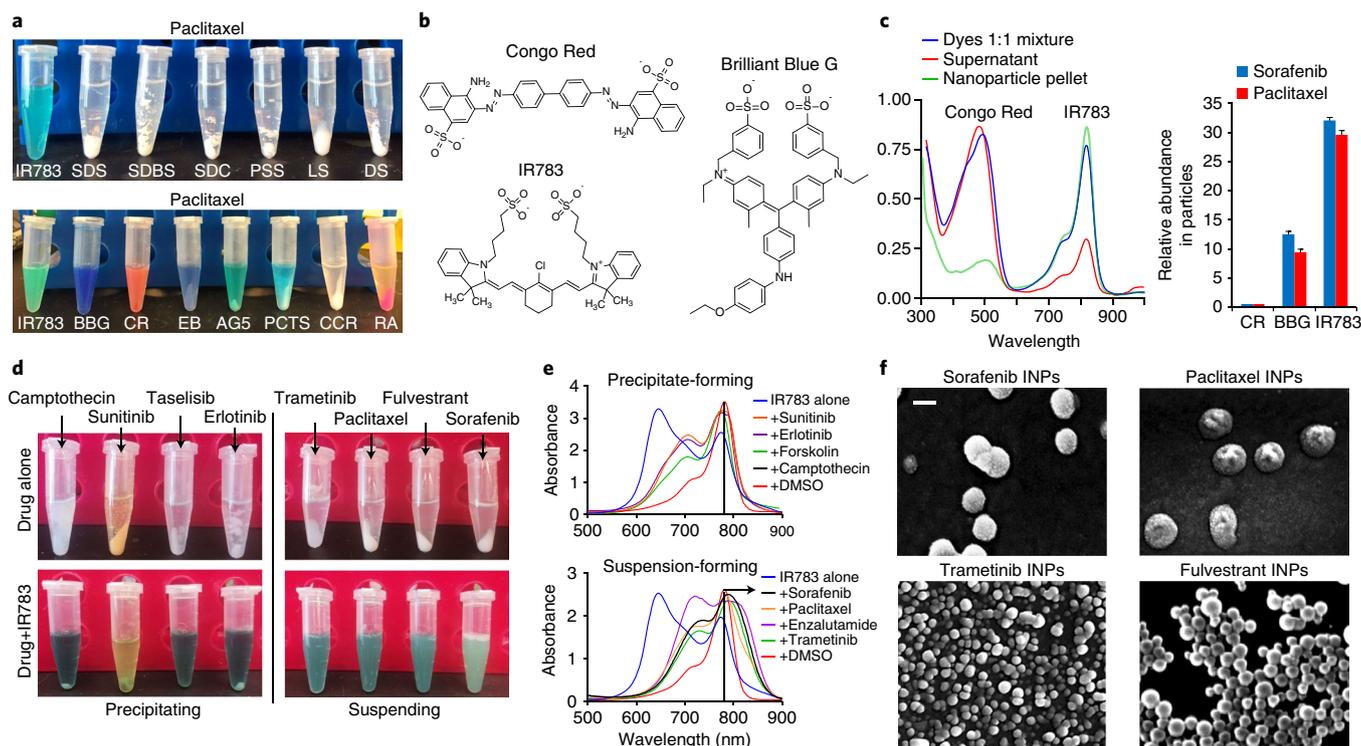


Fig. 1 | Indocyanine-drug self-assembled nanoparticles. **a**, Result of attempted water dispersion of paclitaxel with a panel of excipients. Sodium dodecyl sulfate (SDS), sodium dodecylbenzene sulfonate (SDBS), sodium deoxycholate (SDC), poly-4-styrenesulfonate (PSS), lignin sulfonate (LS) and dextran sulfate (DS), Brilliant Blue G (BBG), Congo Red (CR), Evans Blue (EB), acid green 5 (AG5), phthalocyanine tetra sulfonate (PCTS), Rhodamine 6G (RA), chromoxane cyanine R (CCR). **b**, Chemical structures of dye excipients that most efficiently suspended drugs. **c**, Absorbance spectrum of Congo Red/IR783 dye mixtures before and after suspension of paclitaxel as well as relative abundance of each dye in the suspensions. Error bars are \pm SD (standard deviation) of the mean. **d**, Images of precipitating and suspending IR783–drug mixtures. **e**, Absorption spectra of IR783 upon introduction of drugs which resulted in precipitate formation, and upon introduction of drugs which resulted in suspensions. **f**, Scanning electron microscopy (SEM) images of indocyanine–drug nanoparticles. Scale bar = 100 nm.

Fig. 1b). Of these, the indocyanine IR783 most effectively stabilized drugs, resulting in the complete suspension of five out of nine molecules, and relatively small, monodisperse particles (Supplementary Figs. 1b–d, 2).

Dye mixtures were assessed to determine which dyes were most responsible for nanoparticle formation. Nanoprecipitations of sorafenib and paclitaxel with equimolar dye concentrations were found to contain 30-fold more IR783 than Congo Red, and 3-fold more IR783 than Brilliant Blue G, indicating a preference for indocyanine (Fig. 1c).

We noted distinct colour changes of the IR783–drug mixtures, depending on whether a given drug was successfully suspended (Fig. 1d). Absorbance spectroscopy (Fig. 1e) revealed a relative increase of the IR783 λ_{\max} = 780 nm peak and a decrease of the λ_{\max} = 640 nm peak, consistent with the dissolution of indocyanine H-aggregates^{27,28}. The 780 nm peak also red-shifted up to 850 nm, suggesting the formation of J-aggregates²⁹.

The resulting drug suspensions were characterized by dynamic light scattering (DLS), scanning electron microscopy (SEM), atomic force microscopy (AFM), and transmission electron microscopy (TEM), confirming nanoparticle formation (Fig. 1f, Supplementary Fig. 3). The drug loading was remarkably high, reaching 90% (mass %) for fulvestrant-based INPs, 84% for paclitaxel INPs, 82% for trametinib INPs, and 86% for sorafenib INPs. Additional characterization data is included in the Supplementary Information (Supplementary Fig. 3). Although the high drug content of these particles resembles that of drug nanocrystals, there are distinct differences, including the lack of high-energy processes in the INP

synthesis methods, and particle stability in aqueous solutions free of detergents²³.

Retrospective quantitative structure–property relationships

We assessed a new library of 16 drug compounds, resulting in eight compounds forming nanoparticles with the indocyanine IR783, while the remainder did not (Fig. 2a, Supplementary Fig. 4). As the compounds were mostly similar in molecular weight, hydrophobicity and charge, we investigated a larger set of chemical properties to understand the factors mediating nanoparticle self-assembly. We employed retrospective QSPR analysis to search for molecular descriptors of drugs which correlated with the successful suspension into nanoparticles. We built a training set of the 16 hydrophobic drug molecules (water solubility $< 1 \text{ mg ml}^{-1}$) using a binary classification to denote nanoparticle-forming (ranked as 1) and precipitating compounds (ranked as 0), based on the observed stability of the suspensions. We used Dragon 6 software (*Talette*) to calculate 4,886 descriptors from the molecular structures of the drugs in the training set, which we assessed for correlations with the experimental data of nanoparticle formation (Methods).

The retrospective analysis identified four molecular descriptors that correlated highly with the experimental data for nanoparticle formation, giving Pearson correlation coefficients (r) of greater than 0.85 (Supplementary Fig. 5). One descriptor, SpMAX4_Bh(s), the fourth leading eigenvalue of the Burden matrix, a topochemical index which scores molecules by their geometrical complexity, bond order, and heteroatoms, weighted by the intrinsic state³⁰, gave an impressive Pearson correlation coefficient of $r = 0.98$

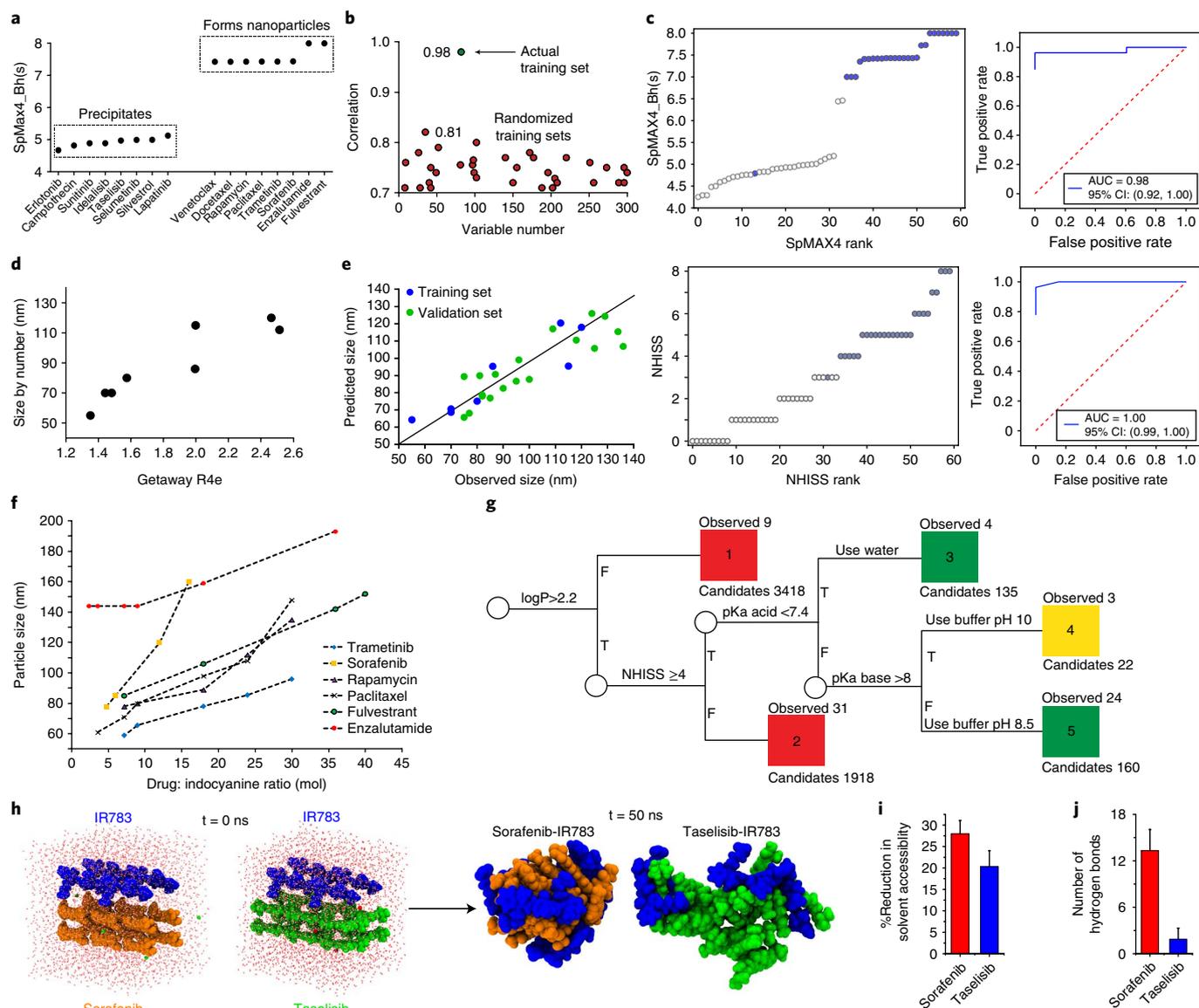


Fig. 2 | Computational prediction and analyses of indocyanine nanoparticle formation. **a**, Training set of 16 drugs experimentally determined to precipitate or form nanoparticles with indocyanine and plotted according to the SpMAX4_Bh(s) descriptor. **b**, Pearson coefficient of SpMAX4_Bh(s) correlation to experimental data for 300 randomized training sets (red circles) and the actual training set (green circles). **c**, Statistical analyses of experimentally validated predictions made with SpMAX4_Bh(s) and NHISS. Descriptor score of each drug plotted by ascending rank of that drug and corresponding receiver-operating characteristic (ROC) curves. Blue = forms nanoparticles experimentally. White = does not form nanoparticles. **d**, A training set of experimentally determined nanoparticle size formed with indocyanine, and linearly correlated to the molecular descriptor GETAWAY R4e. $R^2 = 0.84$ (95% CI [0.22, 0.98], $N = 8$) **e**, Validation set of drug molecules based on the GETAWAY R4e descriptor. **f**, Nanoparticle size as a function of drug:indocyanine ratio. **g**, Decision tree for INP self-assembly. Green = formation of stable nanoparticles in PBS pH 7.4, yellow = formation of stable nanoparticles in basic pH, and red = does not form nanoparticles. Descriptors were calculated by Dragon and ChemAxon. **h**, Snapshots from the top clusters acquired from all-atom molecular dynamics simulations of drug-indocyanine systems. Indocyanine, sorafenib and taselelisib molecules are blue, orange and green, respectively. **i**, Reduction in solvent accessible surface area of drug due to the presence of indocyanine. **j**, Number of intra-nanoparticle hydrogen bonds averaged over the simulated trajectory. Error bars indicate standard deviation, $n = 5,000$. Bar graphs are of mean \pm SD.

(95% confidence intervals (CI) [0.94, 0.99], Methods). The analysis showed that the calculated eigenvalues of the nanoparticle-forming drugs were greater than 7, while the non-assembling drug eigenvalues were less than 5.5 (Fig. 2a). To assess the robustness of this correlation, we compared the training set to 300 randomly generated training sets comprised of the same 16 molecules with scrambled categorization. From these, 80% of the random sets had r values less than 0.7, and the other 20% were less than 0.8; no randomly-generated training sets had an r value near 0.98 (Fig. 2b).

Prediction of nanoparticle self-assembly

To assess the predictive utility of QSNAP Model 1 via the SpMAX4_Bh(s) descriptor, we calculated SpMAX4_Bh(s) values of 400 hydrophobic drugs (defined as having an ALogP2 value, a molecular descriptor of hydrophobicity, of greater than 4.5). Out of the analysed molecules, we identified 131 compounds with an SpMAX4_Bh(s) value of 6.99 and above. A prospective validation set of 44 molecules was constructed by selecting 19 compounds with SpMAX4_Bh(s) values above 6.99 (Supplementary Fig. 6) and 25 compounds with

SpMAX4_Bh(s) values below 6.99 (Supplementary Fig. 7). The validation set was experimentally tested for nanoparticle formation. Remarkably, all drugs but one (avasimibe) behaved as predicted using SpMAX4_Bh(s), while descriptors for hydrophobicity performed only slightly better than random (Fig. 2c, Supplementary Figs. 8, 9 and Supplementary Table 1). It is worth noting that two INPs (ABT737 and venetoclax) exhibited low stability in PBS at pH 7.4. This could be explained by the interaction of basic groups in these molecules ($pK_a \geq 8$, calculated by ChemAxon), with the sulfate groups on IR783, resulting in destabilization of the nanoparticle.

Mechanistic analysis of predictive QSNAP descriptors

It was apparent that compounds with high SpMAX4_Bh(s) values often contained multiple fluorine atoms; therefore, we compared similar drugs with and without fluorine. Our structural screening revealed two similar molecules, celecoxib, with three fluorine atoms and SpMAX4_Bh(s)=7.7, and valdecoxib, with zero fluorines and SpMAX4_Bh(s)=4.7. The drugs behaved experimentally as predicted, where the high-SpMAX4_Bh(s) compound formed nanoparticles but the low-SpMAX4_Bh(s) molecule did not (Supplementary Fig. 10).

We noticed that all four descriptors that correlated highly with nanoparticle formation were weighted by the intrinsic state, so we investigated further. The intrinsic state (I) encodes the electronic state of an atom from the molecular graph as follows:

$$I_i = \frac{(2/L_i)^2 \times \delta_i^v + 1}{\delta_i} \quad (1)$$

where L is the principal quantum number of the atom, δ^v is the number of valence electrons, and δ is the number of sigma electrons³¹. The calculated intrinsic state values for various atoms show that, indeed, fluorine and groups containing double bonds to oxygen (including carbonyl, sulfoxide, nitro, and so on) exhibit the highest intrinsic state values (8 and 7, respectively, Supplementary Table 2). Further investigation of SpMAX4_Bh(s) reveals that all molecular structures with SpMAX4_Bh(s) values ≥ 6.99 contain at least four atoms with an intrinsic state ≥ 7 . Likewise, structures with SpMAX5_Bh(s) ≥ 6.99 have at least five atoms with intrinsic states ≥ 7 , and so on (Supplementary Fig. 11).

Simplified QSNAP descriptors for nanoparticle formation

To simplify the prediction process, we created a new descriptor, Number of High Intrinsic State Substructures (NHISS), calculated as the total number of functional groups in a molecule with intrinsic state atoms ≥ 7 (Methods). Comparing NHISS to SpMAX4_Bh(s) for 430 compounds resulted in an agreement of 98.8% (Supplementary Tables 3 and 4). We thus conclude that NHISS can replace SpMAX4_Bh(s) to assess nanoparticle formation (Fig. 2g).

Prediction of nanoparticle size

We additionally employed QSNAP to predict nanoparticle size. We generated a training set of eight INP-forming drug molecules by measuring particle size by DLS (Supplementary Table 5 and Methods). From the 4,886 descriptors calculated by Dragon 6, we found the GETAWAY R4e correlated strongly with the DLS data; $R^2=0.84$ (95% CI [0.22, 0.98], $N=8$, Fig. 2d). This descriptor is based on the leverage matrix from the spatial coordinates of a molecule using molecular weighting of electronegativity³². We then generated a validation set by calculating this descriptor for an additional 18 nanoparticle-forming drugs and measured the INP sizes experimentally. The resulting nanoparticle sizes were successfully predicted by the GETAWAY R4e descriptor within an error of ± 15 nm (Fig. 2e, Supplementary Table 5). To reduce nanoparticle

particle sizes to below 100 nm, an optimal dye:drug ratio of 1:5 was found (Fig. 2f).

INP assembly prediction of a drug library

We used our predictive model of nanoparticle formation to assess which drugs may form nanoparticles with IR783. We first built a decision tree¹¹, to categorize small molecule drugs from the DrugBank database³³ according to their ability to self-assemble and likely preparation conditions, based on our new understanding of the chemical properties leading to nanoparticle formation from QSNAP Model 1 and experimental observations of pH-related stability issues (Fig. 2g). The decision tree incorporates descriptors and chemical features required for nanoparticle formation. Out of 5,653, the analysis predicted that 317 drugs are capable of forming indocyanine nanoparticles and that 295 drugs out of the 317 would be stable in buffer at pH 7.4 (Methods). The category placements of 63 drug molecules in the decision tree were validated experimentally (Supplementary Table 6).

Molecular simulations of nanoparticle formation

To better understand the self-assembly process, we conducted all-atom replica exchange molecular dynamics (REMD) simulations^{17, 34, 35}. Four indocyanine (IR783) molecules and twelve drug molecules (either sorafenib, a nanoparticle-forming drug, or tasisib, a non-particle-forming drug) were included in a box with explicit water and run with 32 different temperature replicas for 50 ns (Supplementary Fig. 12a,b, additional details in Methods). The simulations resulted in the formation of indocyanine-drug assemblies with clear morphological differences (Fig. 2h). The sorafenib–indocyanine simulation formed a significantly more compact assembly than the tasisib–indocyanine simulation. In the tasisib simulation, the resulting complex remained relatively loosely assembled, as evident from the relative radial particle density (Supplementary Fig. 12c) and solvent accessibility to the drug molecules (Fig. 2i). One possible clue to these differences may be the in the number of hydrogen bonds formed in the two simulations. The sorafenib–indocyanine complex formed over four times as many hydrogen bonds compared to the tasisib–indocyanine complex ($P < 0.001$, Fig. 2j), mainly between the drug molecules themselves (Supplementary Fig. 12f).

Kinase inhibitor-encapsulated INPs

Two INPs encapsulating kinase inhibitors were prepared with the indocyanine IR783 for targeting and anti-tumour efficacy studies. Nanoparticles encapsulating sorafenib, a multikinase inhibitor, and trametinib, a MEK inhibitor, were 80 nm and 55 nm in diameter, and exhibited drug loadings of 86% and 83%, respectively. Both nanoparticles were relatively stable in serum and sink conditions for up to 24 h, and they exhibited sustained drug release for over 48 h (Supplementary Fig. 13).

Mechanism of nanoparticle uptake via endocytosis

The internalization of these nanoparticles was studied in multiple cell lines chosen to represent a range of cell types: endothelial cells, epithelial cancers, leukemia, lymphomas, and fibroblasts. Differential uptake was observed across cell types, with a significant preference for endothelial cells, colon cancer cells, and liver cancer cells (Fig. 3a). We assessed the mechanism of nanoparticle uptake using various inhibitors of endocytic pathways. Inhibitors of caveolin-mediated endocytosis, but not clathrin-mediated endocytosis, significantly attenuated nanoparticle uptake (Fig. 3b). As it has been reported that nanoparticles coated with highly sulfated and aromatic polystyrene sulfonate exhibited differential uptake in cells via caveolae³⁶, we hypothesized that the INPs, incorporating a high loading of hydrophobic drug and sparsely coated by sulfated indocyanines, may elicit similar caveolae-targeting effects.

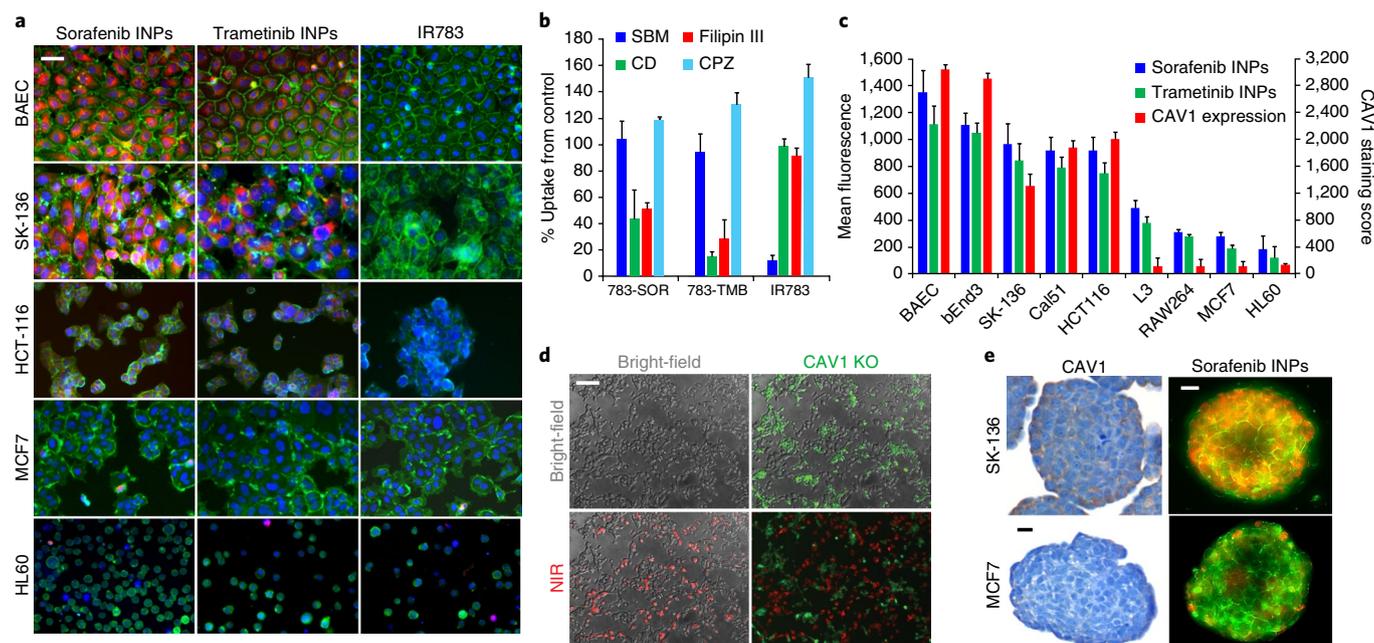


Fig. 3 | Internalization of indocyanine nanoparticles in 2D and 3D cell culture. **a**, Fluorescence micrographs of INP internalization in different cell lines. Green = membrane stain (CellMask), red = indocyanine nanoparticle, blue = nuclear staining. **b**, Inhibition of internalization mechanisms with chemical inhibitors, including cyclodextrin (CD) and filipin III inhibitors of caveolae, chlorpromazine (CBZ) inhibitor of clathrin-mediated endocytosis, and bromo-sulphophthalein (SBM) inhibitor of OAT1-3. **c**, Indocyanine nanoparticle uptake in cell lines, quantified by fluorescence intensity correlated with CAV1 expression ($R^2 = 0.86$). **d**, Nanoparticle (sorafenib INP) uptake in a co-culture of CAV1 knockout (green) and WT HCT116 cells (unstained). Red = near-infrared (NIR) dye fluorescence. **e**, CAV1 staining in tumour spheroids composed of two different cell lines and nanoparticle fluorescence in tumour spheroids. Green = CellMask, red = nanoparticle fluorescence. Scale bar = 25 μm . Bar graphs are of mean \pm SD.

To further investigate the caveolae-targeting hypothesis, we interrogated the Human Protein Atlas and Broad Institute Cancer Cell line Encyclopedia (CCLE) databases for expression levels of CAV1, the main protein scaffold of caveolae (Supplementary Fig. 14a). We found a strong correlation ($R^2 = 0.86$) between CAV1 expression, assessed by immunohistochemistry (Supplementary Fig. 14b), and the nanoparticle fluorescent signal across cell lines (Fig. 3c). We also knocked out the CAV1 gene in a colon cancer cell line, HCT116, wherein GFP was also introduced in the vector. We found that the CAV1 knockout exhibited reduced nanoparticle (sorafenib INP) uptake as compared to the wild-type (WT) cells in a co-culture of the two lines (Fig. 3d). This difference did not depend on the presence of serum or albumin (Supplementary Figs. 15, 16).

We assessed the ability of the nanoparticles to target three-dimensional tumour spheres in vitro using MCF7 and SK-136 (ref. 37) cells in low-adhesion plates (Fig. 3e). Fluorescent imaging of the tumour spheres after 40 min of incubation with $20 \mu\text{gml}^{-1}$ of sorafenib INPs revealed a similar pattern to the two-dimensional (2D) cell culture experiments, preferentially targeting the CAV1-expressing SK-136 tumour spheres (Fig. 3e and Supplementary Figs. 17, 18) despite greater interstitial spacing in SK-136 spheres³⁷.

Pharmacokinetics and biodistribution in healthy mice

We measured the biodistribution of the indocyanine nanoparticles in healthy mice. After intravenous administration by tail vein injection, the nanoparticles appeared in the liver first—within 20 min. The near-infrared signal from the indocyanine in the lungs increased from 0–24 h (Supplementary Fig. 19). The significant lung accumulation may be explained by the large number of endothelial cells and caveolae present in these organs^{38,39}. At 24 h, a low signal in the liver and spleen was observed. The pharmacokinetics of sorafenib, delivered by INPs and in the free drug form, was measured in healthy mice. The blood half-life of sorafenib INPs was

found to be five times longer than that of Kolliphor EL-suspended sorafenib intravenous (i.v.) injection (Supplementary Fig. 19d).

Biodistribution of nanoparticles in hepatic tumour model

Next, we evaluated biodistribution of INPs in a MYC-driven murine hepatic tumour model^{40,41}. To generate autochthonous liver tumours, Sleeping Beauty transposon vectors encoding c-Myc and mutant β -catenin (linked to GFP) were hydrodynamically delivered into immunocompetent FVB mice along with a vector encoding Sleeping Beauty transposase (see Methods). At three weeks after inoculation, tumour nodules could be detected in the liver (Supplementary Fig. 20a–c), and antibody stains of CAV1 revealed its presence in virtually all tumour vessels as well as larger vessels of normal liver tissue (Fig. 4a, Supplementary Fig. 20d,e). Fluorescence imaging of livers harvested 24 h after i.v. injection of sorafenib-encapsulated indocyanine nanoparticles revealed localization specifically within tumour tissue, indicated by co-localization of sorafenib INPs and GFP-positive areas (Fig. 4b). However, we did not observe significant accumulation of sorafenib INPs in normal liver, and free indocyanine did not accumulate in tumours (Supplementary Fig. 20f). Microdistribution of the nanoparticles was imaged in frozen sections and revealed both endothelial (stained with CD31) and cancer cell uptake of the nanoparticles (Fig. 4c, Supplementary Fig. 21). In addition, sorafenib INPs exhibited little accumulation in liver metastases of uveal melanoma tumours, which lack CAV1 (Supplementary Fig. 22), suggesting that CAV1 expression mediates specific uptake of sorafenib INPs in vivo.

Efficacy of INPs in hepatic tumours

To test the therapeutic potential of the INPs in vivo, we compared the anti-tumour efficacy of equivalent drug doses in the murine liver cancer model described above, administered intravenously via INPs or orally for the free drug. Sorafenib or sorafenib INPs

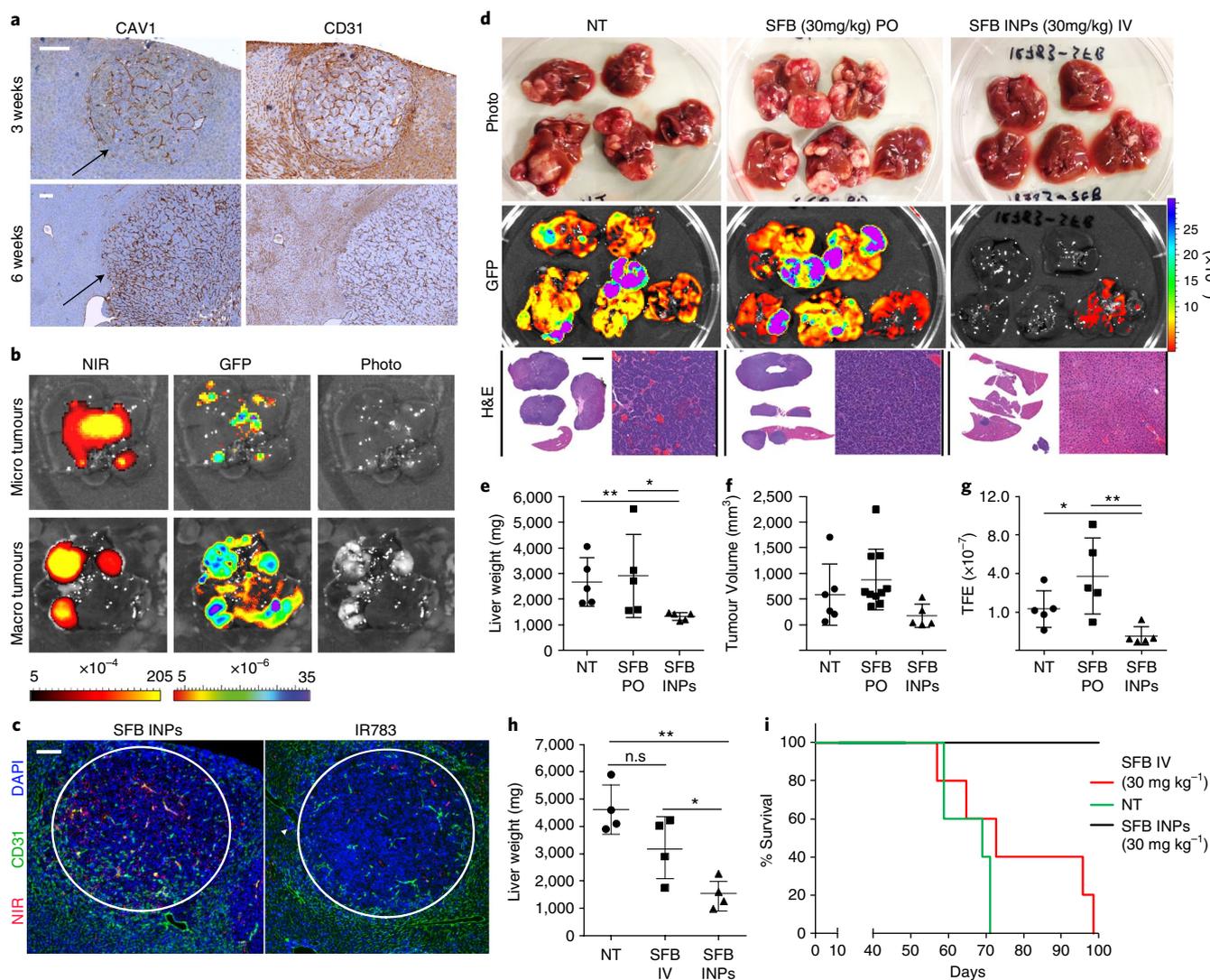


Fig. 4 | Indocyanine nanoparticle targeting and efficacy in MYC-driven autochthonous murine hepatic tumour model. **a**, CAV1 and CD31 staining in liver sections 3 weeks and 6 weeks after hydrodynamic injection. Arrow indicates tumour nodule. Scale bar = 50 μm . **b**, Fluorescence images of livers with multiple GFP-positive tumour nodules 24 h after administration of nanoparticles. NIR = INP indocyanine emission, GFP = cancer fluorescence. **c**, Immunofluorescence images of tissue slices from the autochthonous liver cancer model 24 h after injections of sorafenib INPs or free IR783 dye. Red = NIR fluorescence, green = CD31 antibody for blood vessels, blue = DAPI nuclear stain. Scale bar = 100 μm . **d**, Imaging data including photographs, tumour GFP fluorescence and H&E of livers extracted from mice treated with sorafenib (SFB) orally (PO) or sorafenib INPs (SFB INPs) intravenously (IV) for 28 days. White scale bar = 50 μm , Black scale bar = 10 mm. **e**, Liver weights, $N = 5$, $^{**}P = 0.006$, $^{*}P = 0.0426$. **f**, Tumour volume as measured in the livers if detectable. **g**, Quantification of GFP fluorescence, $N = 5$, $^{**}P = 0.0098$, $^{*}P = 0.0201$. **h**, Comparison of liver weights from mice inoculated with 2x plasmids and treated with IV-administered sorafenib (SFB) or sorafenib INPs (SFB INPs) weekly for three weeks, $N = 4$, $^{**}P = 0.0021$, $^{*}P = 0.039$ $^{ns}P = 0.067$. **i**, Survival data after treatments with i.v.-administered sorafenib (SFB IV) or sorafenib INPs, $N = 5$, log-rank test $z = 3.18$, $P = 0.00113$. All error bars are \pm SD of the mean.

were injected weekly over the course of three weeks. Strikingly, whereas mice treated with free sorafenib exhibited multiple liver tumours at the experimental endpoint of 60 days, the livers of mice treated with nanoparticles containing the same sorafenib concentration showed virtually no residual tumour tissue, evident by visual inspection, GFP imaging, and histology (Fig. 4d). Furthermore, quantitative measures of liver weight, tumour volume, and GFP emission confirmed virtually complete tumour eradication in INP-treated livers (Fig. 4e–g). The INP-delivered sorafenib was also more efficacious in comparison to the free drug administered via the same route in a more aggressive version of the model (Methods, Fig. 4h). A survival study found that 100% of mice administered sorafenib INPs survived, whereas no mice administered the free drug survived (Fig. 4i).

INP-mediated delivery of a MEK inhibitor

We investigated the biodistribution and anti-tumour efficacy of INPs encapsulating the MEK inhibitor trametinib in a CAV1-expressing colon cancer model which is sensitive to MEK inhibition. The subcutaneous HCT116 human colorectal carcinoma model expresses CAV1 in cancer cells and tumour-associated endothelium (Fig. 5a). We assessed the nanoparticle biodistribution in subcutaneous HCT116 xenografts and observed substantial nanoparticle accumulation in tumours (Supplementary Fig. 23). Nanoparticle fluorescence in the tumour tissue was 5-fold higher than in the lungs after 24 h.

We compared the anti-tumour efficacy of trametinib INPs to free trametinib. As with sorafenib, a weekly dose of trametinib did not affect tumour progression, but the nanoparticles, encapsulating

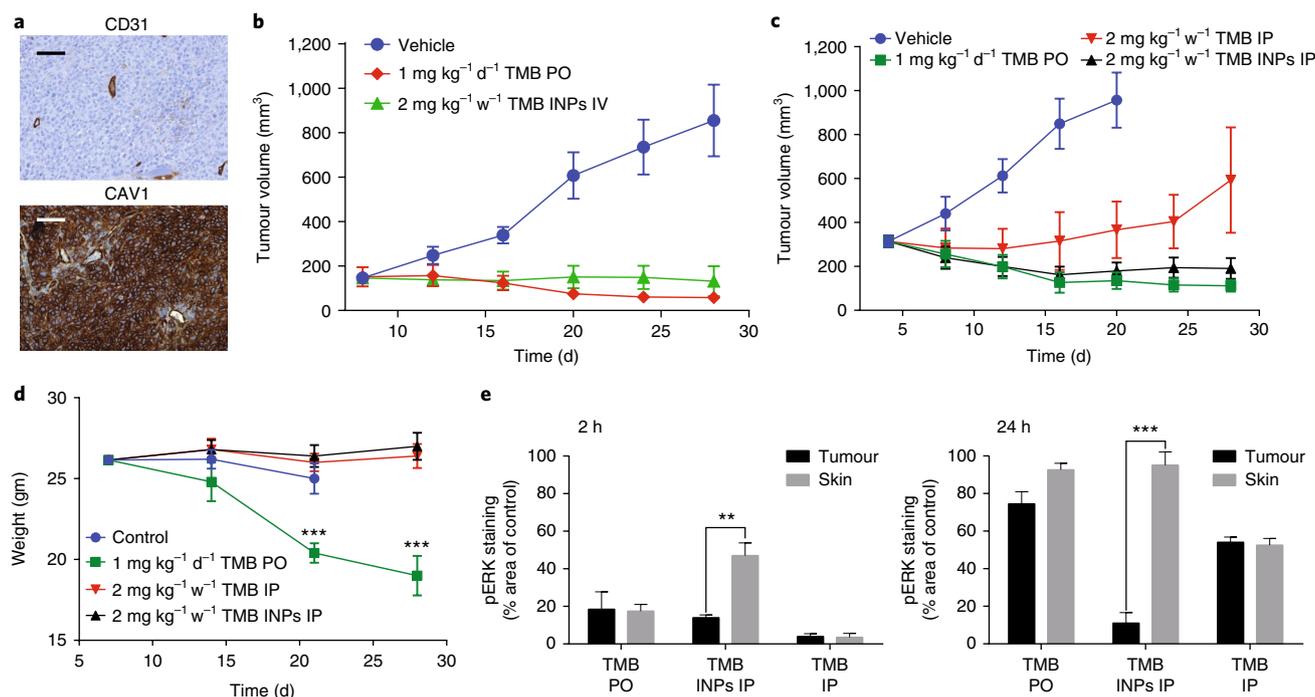


Fig. 5 | Anti-tumour efficacy in HCT116 colon cancer model. **a**, Immunohistochemical staining of tumour section for CAV1 and CD31 expression in HCT116 xenografts two weeks after inoculation. **b**, Tumour growth inhibition in response to i.v. injected nanoparticles weekly or free drug given orally daily ($N=6$) at equivalent doses. **c**, Tumour growth inhibition in response to a weekly dose of nanoparticles or free drug injected intraperitoneally (i.p.; $N=6$) versus an oral daily dose of free drug. Error bars are \pm SE (standard error). **d**, Weights of tumour-bearing mice during treatments. Error bars are \pm SD of mean, $N=6$, $***P=0.0002$, 0.00061 . **e**, Quantification of IHC staining area of pERK in skin and tumour tissue divided by the average staining area of untreated tumours. Error bars are \pm SE of mean, $N=5$, $**P=0.008$, and $***P=0.00014$.

an equivalent dose of drug, elicited significant tumour inhibition (Supplementary Fig. 23d). The anti-tumour efficacy using a weekly dose of 2 mg kg^{-1} trametinib INPs was comparable to a daily dose of 1 mg kg^{-1} free trametinib orally, which was cumulatively 2.5 times higher (Fig. 5b). The trametinib INPs, administered weekly ($2 \text{ mg kg}^{-1} \text{ w}^{-1}$), were as effective as a daily dose of free drug ($1 \text{ mg kg}^{-1} \text{ d}^{-1}$), whereas the weekly dose of free trametinib, administered using the same route of administration, resulted in little tumour inhibition (Fig. 5c).

Toxicity and assessment of off-target MEK inhibition

The toxicities of the INPs and kinase inhibitors were evaluated. Significant mouse weight loss was observed under daily treatments of free trametinib, but no weight loss was apparent for any of the weekly treatments, including with trametinib INPs (Fig. 5d, Supplementary Fig. 23d). We used the downstream phosphorylation status of ERK as a marker for drug activity⁷. The systemic administration of free trametinib resulted in a pronounced inhibition in ERK phosphorylation in the skin and tumour at 2 h, but pERK returned in both after 24 h. In contrast, trametinib INPs elicited prolonged pERK inhibition in the tumour, including after 24 h, but minimal inhibition in the skin was apparent at that time-point (Fig. 5e). As pERK suppression in the skin is associated with dose-limiting skin rash in humans, these results suggest that the nanoparticles could increase the therapeutic index of MEK inhibitors. Finally, drug/dye distribution ratio in tumours and organs was evaluated by comparing the administration of nanoparticles to free drug and dye in HCT116 xenografts (Supplementary Fig. 24a–d). The nanoparticle formulation showed a similar distribution trend of drug and dye in tumours and organs, while the free molecules distributed differently, with up to a 5-fold ratio difference between the tumours, liver and lungs (Supplementary Fig. 24e,f).

Conclusion

This study found a targeted drug carrier nanoparticle, formed by the self-assembly of small molecules, and that the particle self-assembly process can be predicted with an unprecedented degree of certainty via computational methods. The nanoparticles exhibited extremely high drug loadings of up to 90%. Nanoparticles incorporating the kinase inhibitors sorafenib and trametinib selectively targeted CAV1-expressing human colon cancer and autochthonous liver cancer models to yield striking therapeutic effects while preventing inhibition in healthy tissues that normally cause dose-limiting toxicities. The possibility to predict targeted nanoparticle synthesis using molecular structure information of drug compounds presents a significant advancement in the field of drug delivery by facilitating a process that is normally conducted by bench chemistry.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41563-017-0007-z>.

Received: 23 January 2016; Accepted: 4 December 2017;

Published online: 05 February 2018

References

- Peer, D. et al. Nanocarriers as an emerging platform for cancer therapy. *Nat. Nanotech.* **2**, 751–760 (2007).
- Schroeder, A. et al. Treating metastatic cancer with nanotechnology. *Nat. Rev. Cancer* **12**, 39–50 (2012).
- Yaari, Z. et al. Theranostic barcoded nanoparticles for personalized cancer medicine. *Nat. Commun.* **7**, 13325 (2016).
- Wilhelm, S. et al. Analysis of nanoparticle delivery to tumours. *Nat. Rev. Mater.* **1**, 16014 (2016).

5. Cheng, Z., Al Zaki, A., Hui, J. Z., Muzykantov, V. R. & Tsourkas, A. Multifunctional nanoparticles: Cost versus benefit of adding targeting and imaging capabilities. *Science* **338**, 903–910 (2012).
6. Lammers, T., Kiessling, F., Hennink, W. E. & Storm, G. Drug targeting to tumors: Principles, pitfalls and (pre-) clinical progress. *J. Control. Release* **161**, 175–187 (2012).
7. Shamay, Y. et al. P-Selectin is a nanotherapeutic delivery target in the tumor microenvironment. *Sci. Transl. Med.* **8**, 345ra87 (2016).
8. Mizrachi, A. et al. Tumour-specific Pi3k inhibition via nanoparticle-targeted delivery in head and neck squamous cell carcinoma. *Nat. Commun.* **8**, 14292 (2017).
9. Maojo, V. et al. Nanoinformatics: A new area of research in nanomedicine. *Int. J. Nanomed.* **7**, 3867–3890 (2012).
10. Irwin, J. J. et al. An aggregation advisor for ligand discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).
11. Seidler, J., McGovern, S. L., Doman, T. N. & Shoichet, B. K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **46**, 4477–4486 (2003).
12. Feng, B. Y., Shelat, A., Doman, T. N., Guy, R. K. & Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **1**, 146–148 (2005).
13. Alskar, L. C., Porter, C. J. & Bergstrom, C. A. Tools for early prediction of drug loading in lipid-based formulations. *Mol. Pharm.* **13**, 251–261 (2016).
14. Fourches, D. et al. Quantitative nanostructure-activity relationship modeling. *ACS Nano* **4**, 5703–5712 (2010).
15. Puzyn, T. et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat. Nanotech.* **6**, 175–178 (2011).
16. Zhang, Y. et al. Lipid-modified aminoglycoside derivatives for in vivo siRNA delivery. *Adv. Mater.* **25**, 4641–4645 (2013).
17. Roxbury, D., Jagota, A. & Mittal, J. Sequence-specific self-stitching motif of short single-stranded DNA on a single-walled carbon nanotube. *J. Am. Chem. Soc.* **133**, 13545–13550 (2011).
18. Lee, O. S., Stupp, S. I. & Schatz, G. C. Atomistic molecular dynamics simulations of peptide amphiphile self-assembly into cylindrical nanofibers. *J. Am. Chem. Soc.* **133**, 3677–3683 (2011).
19. Frederix, P. W. et al. Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. *Nat. Chem.* **7**, 30–37 (2015).
20. Shi, C. et al. Drug-specific nanocarrier design for efficient anticancer therapy. *Nat. Commun.* **6**, 7449 (2015).
21. Alphandery, E., Grand-Dewyse, P., Lefevre, R., Mandawala, C. & Durand-Dubief, M. Cancer therapy using nanoformulated substances: Scientific, regulatory and financial Aspects. *Expert. Rev. Anticancer. Ther.* **15**, 1233–1255 (2015).
22. Agarwal, A., Lvov, Y., Sawant, R. & Torchilin, V. Stable nanocolloids of poorly soluble drugs with high drug content prepared using the combination of sonication and layer-by-layer technology. *J. Control. Release* **128**, 255–260 (2008).
23. Muller, R. H. & Keck, C. M. Challenges and solutions for the delivery of biotech drugs—a review of drug nanocrystal technology and lipid nanoparticles. *J. Biotech.* **113**, 151–170 (2004).
24. McLaughlin, C. K. et al. Stable colloidal drug aggregates catch and release active enzymes. *ACS Chem. Biol.* **11**, 992–1000 (2016).
25. Shi, C., Wu, J. B. & Pan, D. Review on near-infrared heptamethine cyanine dyes as theranostic agents for tumor imaging, targeting, and photodynamic therapy. *J. Biomed. Opt.* **21**, 50901 (2016).
26. Yang, X. et al. Near Ir heptamethine cyanine dye-mediated cancer imaging. *Clin. Cancer Res.* **16**, 2833–2844 (2010).
27. McArthur, E. A., Godbe, J. M., Tice, D. B. & Weiss, E. A. A study of the binding of cyanine dyes to colloidal quantum dots using spectral signatures of dye aggregation. *J. Phys. Chem.* **116**, 6136–6142 (2012).
28. Slavnova, T. D., Gorner, H. & Chibisov, A. K. Cyanine-based J-aggregates as a chirality-sensing supramolecular system. *J. Phys. Chem. B* **115**, 3379–3384 (2011).
29. Fofang, N. T., Grady, N. K., Fan, Z., Govorov, A. O. & Halas, N. J. Plexciton dynamics: exciton-plasmon coupling in a J-aggregate–Au nanoshell complex provides a mechanism for nonlinearity. *Nano Lett.* **11**, 1556–1560 (2011).
30. Todeschini, R. & Consonni, V. *Handbook of Molecular Descriptors* Vol. 11 (eds Mannhold, R., Kubinyi, H. & Timmerman, H.) 285–286 (Wiley, New York, NY, 2008).
31. Kier, L. B. & Hall, L. H. An electrotopological-state index for atoms in molecules. *Pharm. Res.* **7**, 801–807 (1990).
32. Consonni, V., Todeschini, R. & Pavan, M. Structure/response correlations and similarity/diversity analysis by Getaway descriptors. 1. Theory of the novel 3d molecular descriptors. *J. Chem., Inf. Comput. Sci.* **42**, 682–692 (2002).
33. Wishart, D. S. et al. Drugbank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–672 (2006).
34. Rhee, Y. M. & Pande, V. S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **84**, 775–786 (2003).
35. Zhou, R. in *Protein Folding Protocols* (eds Bai, Y. & Nussinov, R.) 205–223 (Springer, New York, NY, 2007).
36. Voigt, J., Christensen, J. & Shastri, V. P. Differential uptake of nanoparticles by endothelial cells through polyelectrolytes with affinity for caveolae. *Proc. Natl Acad. Sci. USA* **111**, 2942–2947 (2014).
37. Jena, P. V. et al. Photoluminescent carbon nanotubes interrogate the permeability of multicellular tumor spheroids. *Carbon* **97**, 99–109 (2016).
38. Wang, Z., Tirupathi, C., Cho, J., Minshall, R. D. & Malik, A. B. Delivery of nanoparticle: Complexed drugs across the vascular endothelial barrier via caveolae. *IUBMB Life* **63**, 659–667 (2011).
39. Chrastina, A., Massey, K. A. & Schnitzer, J. E. Overcoming in vivo barriers to targeted nanodelivery. *WIREs. Nanomed. Nanobiotechnol.* **3**, 421–437 (2011).
40. Chen, X. & Calvisi, D. F. Hydrodynamic transfection for generation of novel mouse models for liver cancer research. *Am. J. Path.* **184**, 912–923 (2014).
41. O'Donnell, K. A. et al. A Sleeping Beauty mutagenesis screen reveals a tumor suppressor role for Nco2/Src-2 in liver cancer. *Proc. Natl Acad. Sci. USA* **109**, E1377–1386 (2012).

Acknowledgements

This work was supported in part by the NIH New Innovator Award (DP2-HD075698), National Cancer Institute (CA 013106), and Cancer Center Support Grant (P30 CA008748), the Expect Miracles Foundation - Financial Services Against Cancer, the Anna Fuller Fund, the Louis V. Gerstner Jr. Young Investigator's Fund, the Frank A. Howard Scholars Program, Cycle for Survival, the Alan and Sandra Gerry Metastasis Research Initiative, Mr. William H. Goodwin and Mrs. Alice Goodwin and the Commonwealth Foundation for Cancer Research, the Experimental Therapeutics Center, the Imaging & Radiation Sciences Program, and the Center for Molecular Imaging and Nanotechnology of Memorial Sloan Kettering Cancer. This work is supported in part by a New York State Department of Health Fixed Term Agreement (Contract# DOH01-C30315GG-3450000). The opinions, results, findings and/or interpretations of data contained therein are the responsibility of the contractor and do not necessarily represent the opinions, interpretations or policy of the state or, if funded with federal funds, the applicable federal funding agency. Y.S. was supported by the Center for Metastasis Research (CMR) Scholars Fellowship Program. D.R. was supported by an American Cancer Society – Roaring Fork Valley Postdoctoral Fellowship. S.W.L. is a Howard Hughes Medical Institute Investigator and the Geoffrey Beene Chair for Cancer Biology (MSKCC). J.B. was supported by the Tow Foundation Postdoctoral Fellowship, Center for Molecular Imaging and Nanotechnology at MSKCC. We would like to thank the following facilities at MSKCC: Molecular Cytology Core Facility, Small Animal Imaging, Anti-tumor Assessment, and Electron Microscopy. The molecular dynamics work used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation grant number TG-MCB-130013. We would like to thank N. Lampen for assistance with electron microscopy. We would also like to thank P.V. Jena, R. Williams, J. Harvey, T. Galassi, J. Humm and C. Horoszko for helpful discussions.

Author Contribution

Y.S. and D.A.H. conceived the project and designed experiments. Y.S. analysed data, designed, and conducted the self-assembly experiments. D.R. performed MD experiments and analysis. D.E.T. and J.L. performed the in vivo liver cancer model. Y.S., V.K.R., A.M. and J.S. performed all other in vivo experiments, electron microscopy and tissue staining. Y.S., V.K.R., K.N., J.L.S., M.R.N., K.C. and J.S. performed in vitro experiments. Y.S. developed the tumour spheroid model. E.B., J.L.S., K.N., R.S., M.R.N., K.C., K.S.G., M.D. and D.C.J. performed experiments for the computational drug screening and nanoparticle self-assembly validation experiments. J.B. performed F-NMR studies for drug biodistribution. M.I. performed self-assembly categorization of DrugBank small molecule drugs. M.I. and J.D.C. conducted statistical analysis of QSNAP descriptors. Y.S. and D.A.H. wrote the paper. D.A.H., S.W.L. and J.D.C. supervised the research.

Competing interests

J. D. C. is a member of the Scientific Advisory Board for Schrödinger, LLC.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41563-017-0007-z>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.A.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Materials and Reagents. All non-drug chemicals were purchased from Sigma Aldrich (St. Louis, MO). Solvents were purchased from Acros Organics (Morris Plains, NJ). All drugs were purchased from LC-Laboratories (Woburn, MA), except glibenclamide, glimepiride and cyclosporine, which were purchased from Sigma Aldrich, TAK632 from AdooQ Bioscience (Irvine, CA), and talazoparib, fulvestrant, venetoclax, selumetinib and taselisib from Selleckchem (Houston, TX).

Cell lines. All cell lines reported in this study except SK-136 (see details below in tumour spheroids section) were purchased from ATCC (Manassas, VA) and were cultured in DMEM (Life Technologies) medium supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin Streptomycin (Life Technologies) in a 37 °C and 5% CO₂ incubator. Cells were checked for mycoplasma contamination by PCR. The MCF7 cell line is on the ICLAC Register of Misidentified Cell Lines (the contaminating cell line is OVCAR-8), and therefore we used these cells only up to ten passages after receiving the frozen vial directly from ATCC. Our CAV1 immunohistochemistry data for MCF7 is in agreement with the Broad Institute Cancer Cell Line Encyclopedia mRNA data, but not with OVCAR-8.

Preparation of indocyanine nanoparticles. We added 0.1 ml of each drug, dissolved in DMSO (10 mg ml⁻¹), dropwise (20 µL per 15 sec) to a 0.6 ml aqueous solution containing IR783 (Sigma Aldrich, 1 mg ml⁻¹) and 0.05 mM sodium bicarbonate or deionized water (for drugs from group 3) under slight vortexing. The solution was centrifuged twice (20,000 g, 30 min), and the pellet was re-suspended in 1 ml of sterile PBS. In cases of a pellet that was difficult to re-suspend, it was ultrasonicated for 5 sec with a 1/8 inch probe tip (Sonics & Materials) at 40% intensity. The nanoparticles were lyophilized in a 5% saline/sucrose solution. Absorbance spectra of solutions and suspensions were acquired using a TECAN M1000 plate reader.

Nanoparticle characterization. Dynamic light scattering (DLS) and zeta potential measurements were conducted using a Zetasizer Nano ZS (Malvern). Scanning electron microscopy (SEM) was conducted using a Zeiss Supra 25 Field Emission scanning electron microscope. SEM samples were prepared by gold sputtering and critical point drying. For atomic force microscopy (AFM) measurements, nanoparticles were observed on a freshly cleaved mica surface using an Asylum MFP 3D Bio AFM with an Olympus AC240TS probe. For transmission electron microscopy (TEM), a carbon-coated copper TEM grid (Ted Pella) was used with a JEOL 1200 EX transmission electron microscope operated at 80 kV.

Quantitative structure-nanoparticle assembly prediction (QSNAP). Training set molecules were selected from drugs with solubilities of under 0.1 mg ml⁻¹ in water. Molecular structure mol2 files, obtained from ChemSpider.com, were energy minimized using the ChemBio3D Ultra 14 Suite. We calculated 4,886 molecular descriptors for each molecular structure using Dragon 6 software (*Talette*). The descriptors were correlated to the binary experimental observations of nanoparticle formation, confirmed by DLS (entered as a rank of 5 in the vector) or precipitation (denoted as a rank of 0) by visual inspection.

QSNAP Model 1: Prediction of INP formation. The training set consisted of 16 drug molecules which were hydrophobic (water solubility <1 mg ml⁻¹) and highly soluble in DMSO (>10 mg ml⁻¹). 4,886 descriptors from Dragon 6 software (*Talette*) were calculated, and their correlation with the experimental data of nanoparticle formation was analysed. The descriptor with highest Pearson correlation coefficient (*r*), SpMAX4_Bh(s), was used for self-assembly prediction. 95% confidence intervals of *r* were estimated based on a procedure described previously⁴². An SpMAX4_Bh(s) value of 6.99 was observed to be the critical value. Molecules with SpMAX4_Bh(s) values equal to or above 6.99 were predicted to form nanoparticles with indocyanine dye. According to this model, 18 predicted nanoparticle formers and 18 predicted nanoparticle non-formers were picked according to commercial availability, to build a validation set. Nanoparticle formation of the validation set was tested under the same experimental conditions.

QSNAP Model 2: Prediction of INP size. A training set was formed using eight INP-forming drug molecules with experimentally observed sizes. Using the Dragon 6 software, the GETAWAY R4e descriptor was found to exhibit the greatest correlation to INP size. We built a linear model using the regression module of Microsoft Excel 2010 Data Analysis ToolPak to predict nanoparticle sizes based on the GETAWAY R4e descriptor as follows:

$$\text{Size} = 48.30(\text{R4e}) - 0.99 \quad (2)$$

Coefficients in the above equation were determined by linear regression to eight data points of the training set. 95% confidence intervals of R² were estimated based on a procedure described in ref. ⁴². The descriptor was calculated for an additional 18 nanoparticle-forming drugs, to be used as a validation set. The nanoparticle sizes were measured experimentally by DLS (Supplementary Table 3).

Molecular dynamics simulations of INP self-assembly. Four indocyanine molecules and twelve drug molecules of sorafenib or taselisib were placed in a 5 nm × 5 nm × 5 nm water box with periodic boundary conditions containing approximately 3,700 TIP3P model water molecules and sodium counter-ions to balance the negative charges of the indocyanine. The total system was comprised of ~12,000 atoms. To run the REMD simulations, the Gromacs 4.6.7 simulation package was used with the Charmm36 force field. Long-range electrostatics were calculated using the particle mesh Ewald method with a 0.9 nm real space cutoff. For van der Waals interactions, a cutoff value of 1.2 nm was used. Simulation parameters for the indocyanine and drug particles were obtained from SwissParam (Swiss Institute of Bioinformatics). The indocyanine-drug configurations were energy minimized and subjected to 100 ps NVT equilibration at 300 K. Thirty-two replicas of the configuration were created with temperatures ranging from 300 K to 563 K. Temperature intervals increased with absolute temperature to maintain uniform exchange probability around 10% acceptance. The 32 replicas were run in parallel for 50 ns of NVT production. Exchange between adjacent temperature replicas was attempted every 2 ps. The time step of the simulation was 2 fs. The trajectories were saved every 10 ps, yielding a total of 5,000 snapshots for production analysis. Structures were visualized in VMD.

Indocyanine nanoparticle self-assembly prediction of DrugBank small molecule drugs. Based on our understanding of the chemical properties leading to drug-dye nanoparticle formation from QSNAP analysis and inspection of 52 drug molecules experimentally determined to form or not form INPs (all drug molecules listed in Table S1, except the following which were not experimentally tested at the time for INP formation: bithionol, probucol, cholesterol, mubritinib, pyrene, ospemifene and tazarotene), a decision tree was built that categorizes small molecule drugs from DrugBank according their ability to self-assemble and their predicted preparation conditions (Fig. 2g). We analysed all small molecule type drug entries with reported structures in DrugBank Version 4.5.0 (<https://www.drugbank.ca/> released on 20 April 2016) and categorized them according to the decision tree.

Simplified Molecular-Input Line-Entry System (SMILES) strings, LogP, strongest acidic pK_a and strongest basic pK_a values were extracted from DrugBank records. LogP, strongest acidic pK_a, and strongest basic pK_a are calculated physical properties reported in DrugBank by ChemAxon. All Python scripts used in the analysis, the dataframe of chemical descriptors and partitioning results based on the decision tree (Fig. 2g) can be found in the *nano-drugbank* repository, hosted on the collaborative open source software development platform GitHub (<http://github.com/choderalab/nano-drugbank>). A full list of small molecule drugs from DrugBank predicted to self-assemble into indocyanine nanoparticles according to the decision tree analysis is available as a CSV file in the repository ([df_drugbank_decision_tree.csv](#)).

Calculation of the Number of High Intrinsic State Substructures (NHISS) descriptor. NHISS was calculated as the total number of functional groups in a molecule with intrinsic state atoms ≥ 7. These groups included fluorine atoms (-F) and double-bonded oxygen atoms (=O). To determine the NHISS scores of a set of molecules, we conducted a substructure search for fluorine, carbonyl, nitro, nitroso, sulfinyl, and sulfonyl functional groups using the SMILES strings of each molecule. A substructure search using SMILES Arbitrary Target Specification (SMARTS) was performed programmatically using the OpenEye OEChem Toolkit. SMARTS used for each functional group and matching substructures patterns are reported in Supplementary Fig. 25. SMARTS editor and SMARTSviewer⁴³ were used to generate SMARTS strings that match desired substructures. Number of substructure matches (*n*) of fluorine, carbonyl, sulfinyl, and nitroso groups incremented the NHISS value by 1, and nitro and sulfone groups incremented by 2.

$$\text{NHISS} = n_{\text{fluorine}} + n_{\text{carbonyl}} + n_{\text{sulfinyl}} + 2n_{\text{sulfonyl}} + n_{\text{nitroso}} + 2n_{\text{nitro}} \quad (3)$$

A performance comparison of drug categorization based on NHISS ≥ 4 and SpMAX4_Bh(s) ≥ 7.0 critical values was analysed and reported in a confusion matrix (Supplementary Table 3, 4). Python scripts used in the calculation of the NHISS descriptor can be found in the *nano-drugbank* GitHub repository at <https://github.com/choderalab/nano-drugbank>.

Analysis of self assembly prediction performance of hydrophobicity descriptors compared to structure-based descriptors SpMAX4_Bh(s) and NHISS. To analyse the discrimination performance of logP (OpenEye XlogP, Chem3D ClogP, Dragon ALOGP2), logD (Chemicalize, pH = 7.4), SpMAX4_Bh(s) and NHISS descriptors, Receiver Operating Characteristic (ROC) curves were plotted, using experimental data points (all molecules listed in Table S1, except pyrene, which was not included in this analysis since it is not a drug) that pass the first decision point of the decision tree: 60 drugs with ChemAxon logP > 2.2 that were tested for nanoparticle formation.

We computed the area under the curve (AUC) for the receiver-operator characteristic (ROC) function, which measures how much information a descriptor carries about whether the drug will form or not form INPs. An ROC AUC of 0.5 indicates that the descriptor carries no information about whether the

drug will form INPs, whereas an AUC of 1.0 indicates the descriptor is perfectly predictive; intermediate values indicate various degrees of predictiveness. This measure is standard in design fields such as virtual screening to assess the relative utility of various predictive models

We computed 95% confidence intervals of the ROC AUC by bootstrapping experimental data: the dataset of 60 drug molecules was resampled 1000 times with replacement. Confidence intervals are estimated based on the distribution of AUC values of resampled datasets of size 60. Python scripts used for the performance comparison of QSNAP Model 1 descriptors (NHIS and SpMAX4_Bh(s)) and various hydrophobicity descriptors are available in the IR783-QSNAP GitHub repository at <https://github.com/choderalab/IR783-QSNAP>.

Drug release measurements. Nanoparticles were incubated in PBS at pH 5.5 and 7.4 at 37 °C or with serum with a concentration equivalent to 1 μM of drug. The amount of released drug was determined by extracting into ethanol/acetonitrile and measuring the absorbance at 260 nm (for sorafenib) and 314 nm for trametinib using a UV-VIS-NIR spectrophotometer (Jasco 670) or plate reader (Tecan Infinite M1000). All experiments were carried out in duplicate.

Protein adsorption. Protein adsorption was evaluated using surface zeta potential measurements of nanoparticles incubated with escalating doses of bovine serum albumin (BSA) (0–40 mg ml^{-1}). Measurements were performed 15 min after incubation with BSA with a Malvern Zetasizer Nano ZS system.

Nanoparticle uptake in cell lines. Cell lines bEnd.3, BAEC, SK-136, L3, MCF7, and HL60 were used. The cells were plated in a 24-well plate (50,000 cells in 1 ml) and incubated with 20 $\mu\text{g ml}^{-1}$ of nanoparticles for 45 min and with CellMask Green (Life Technologies) for another 15 min to stain cell membranes and Hoechst 33342 (Life Technologies) to stain nuclei. The cells were rinsed twice with PBS. Images were acquired with an inverted Olympus IX51 fluorescent microscope equipped with an XM10IR Olympus camera and an Excite Xenon lamp. Similar exposure times and excitation intensities were applied throughout all experiments. Filter sets: cell membrane: ex 488 nm, em 525 nm, nucleus: ex 350 nm, em 460 nm, IR783 dye in particles: ex 780 nm, em 820 nm. Images were processed with ImageJ software.

Development of tumour spheroids. To generate multi-cellular tumour spheroids, we developed a cell line, SK-136, derived from the autochthonous liver cancer model. The cells were generated and harvested from c-MYC/ β -catenin amplified hepatoblastoma cells from FVB mice using a method described previously. The harvested cells were plated on ultralow-attachment 96-well plates (Corning) and incubated for three days. The wells were examined with an inverted light microscope to confirm the formation of multicellular tumour spheroids. The wells containing tumour spheres were centrifuged, trypsinized, and seeded in 75 cm cell-culture-treated flasks with DMEM. This process was repeated three times to generate a sub-clone of spheroid-forming cells. To identify CAV1 expression in 7-day-old tumour spheroids, they were collected at the bottom of an Eppendorf tube, suspended in paraformaldehyde, and embedded in paraffin. 10 μm slices were stained with anti-caveolin-1 antibody (Cell Signaling, cat no. 3267, 1 $\mu\text{g ml}^{-1}$) as well as haematoxylin and eosin (H&E) staining. To characterize the surface of the tumour spheroids, SK-136 cells were grown in ultralow-attachment flasks (Corning) for five days. Once the spheres were formed, the media containing tumour spheres was removed and placed in 1 ml Eppendorf tube. The spheres were allowed to settle by gravity for 2 min and the media was replaced with fresh media. The spheroids were placed on poly-L-lysine-coated plastic coverslips (ThermoFisher). The spheroids were then fixed in 2.5% paraformaldehyde in 0.075 M cacodylate buffer for one hour, rinsed in cacodylate buffer, and dehydrated in a graded series of alcohols: 50%, 75%, 95% and 100%. The samples were then dried in a JCP-1 Critical Point Dryer (Denton). The coverslips were attached to SEM stubs and sputter-coated with gold/palladium using a Desk IV sputter system (Denton Vacuum). The images were obtained in a Scanning Field Emission Supra 25 scanning electron microscope (Zeiss).

Penetration of nanoparticles in tumour spheroids. 10⁴ SK-136 cells were seeded in 25 cm² ultralow attachment flasks (Corning) and grown for seven days in DMEM with media replacement every three days. When spheres reached a diameter of approximately 250 μm , 0.2 ml of growth suspension was plated in normal adhesion 96-well plates, yielding 3–5 spheres per well. After 30 min, spheres adhered to the bottom of the wells. Nanoparticles were added at a concentration of 50 $\mu\text{g ml}^{-1}$ and incubated for 20–40 min. The wells were washed three times with HBSS buffer and imaged with an inverted Olympus IX51 fluorescence microscope equipped with a XM10 Olympus CCD camera. The fluorescence intensity was analysed using ImageJ software.

Animal Studies. All of the animal studies were conducted according to protocols approved by the Memorial Sloan Kettering Cancer Center Institutional Animal Care and Use Committee (IACUC). Mice were maintained and treated in accordance with the institutional guidelines of the Memorial Sloan Kettering Cancer Center.

MYC/ β -catenin driven liver tumour studies. Hydrodynamic transfection of C-MYC and β -catenin was performed as follows: 10 μg pT3-EF1a-c-myc, 10 μg pT3-EF1a- β -CateninT41A-IRES-GFP and CMV-SB13 Sleeping Beauty transposase (1:5 ratio) were mixed in sterile saline solution. A total volume of plasmid-saline solution mix corresponding to 10% of the body weight was injected into the lateral tail vein of 6- to 8-week-old female FVB/N mice (Jackson Laboratory, Maine, USA) within 5–7 seconds. Mice were administered either 30 mg kg^{-1} sorafenib orally, or 30 mg kg^{-1} sorafenib in sorafenib INP form via tail vein injection. For targeting and biodistribution experiments, mice were injected with sorafenib INPs or indocyanine 3 weeks and 6 weeks after transfection. Livers were harvested 24 h after injection. For efficacy studies, treatments were administered weekly for three weeks. Livers were harvested at day 59. Tumours were evaluated using fluorescence imaging (IVIS imaging system, Xenogen Corp., Hopkinton, MA) and immunohistochemistry (H&E). Tumour volume was measured using a caliper. For the aggressive model 20 μg pT3-EF1a-c-myc, 20 μg pT3-EF1a- β -CateninT41A-IRES-GFP and CMV-SB13 Sleeping Beauty transposase were used (double that above).

Colon cancer xenograft studies. Six-week-old female athymic NU/NU nude mice purchased from Charles River Laboratories were injected with 5×10^5 of HCT116 human colorectal carcinoma cells subcutaneously in 100 ml culture media/Matrigel (BD Biosciences) at a 1:5 ratio. Animals were randomized at a tumour volume of 70 to 120 mm³ into four to six groups, with $n = 8$ –10 tumours per group. Animals were treated per os (p.o.) with trametinib (1 mg kg^{-1}) or i.v. with trametinib INPs (1 mg kg^{-1}) once a week. Tumour size was measured with a digital caliper, and tumour volumes were calculated using the formula: (length \times width²) \times ($\pi/6$). Animals were euthanized using CO₂ inhalation. Mice were housed in air-filtered laminar flow cabinets with a 12-hr light/dark cycle and food and water ad libitum.

Blood pharmacokinetics. Serial blood samples were collected in lithium–heparin-containing tubes before and at 0.05, 0.5, 2, 4, 24 and 48 h after the administration of the first dose of sorafenib INPs or free sorafenib in Kolliphor EL. Samples were processed within 30 min of collection by centrifugation for 10 min at 1500 g under refrigeration ($\sim 4^\circ\text{C}$). The resultant plasma was stored at -70°C until subsequent analysis for sorafenib using a validated liquid chromatography/tandem mass spectrometry method. Briefly, sorafenib was extracted from plasma using acetonitrile precipitation. Separation of sorafenib, C18 (150 mm \times 2.1 mm internal diameter, 3.5 μm ; Agilent Technologies USA) analytical column using a mobile phase consisting of acetonitrile containing 0.1% trifluoroacetic acid and deionized water containing 0.1% trifluoroacetic acid. The gradient from 0 to 95% acetonitrile in 17 min and flow at 1 ml min^{-1} showed a retention time of 12.5 min and absorbance at 280 nm.

In vivo imaging of INPs with IVIS. 24 h after tail vein injection of 200 μl (1 mg ml^{-1}) INP (sorafenib, trametinib, fulvestrant and valrubicin). Organs were extracted and placed on a Petri dish (liver, kidney, lungs, heart, spleen and tumour if present). Images were taken with an IVIS imaging system (Xenogen Corp.). Radiance efficiency (photons $\text{sec}^{-1} \text{cm}^{-2}$) was calculated for the tumour region of interest (ROI) using LivingImage V4.2 software.

Drug biodistribution. 24 h after administration of nanoparticles or free drugs, the organs were cut into small pieces, homogenized with lysis buffer, and the drugs mixed with chloroform overnight. The solution mixture was filtered and the homogenate was treated twice with chloroform with repeated filtration. Chloroform was evaporated under vacuum and reconstituted in chloroform-d6. 19F-NMR spectroscopy was applied for fluorine quantification which arises solely from the drugs.

Liver metastasis model of uveal melanoma in NOD SCID GAMMA (NSG mice). Human liver metastatic-enriched uveal melanoma cells expressing GFP-luciferase (L3 from OMM-1 cells, ATCC) were supplied by V. Rajasekhar. 5×10^5 cells were injected via the retro-orbital sinus on NSG mice (JAX Laboratories). The appearance of liver metastases by bioluminescence was observed within 14 days after inoculation. Nanoparticles were injected 24 h before imaging (with an IVIS imaging system, Xenogen Corp., Hopkinton, MA).

Immunohistochemistry. For xenograft samples, dissected tissues were fixed immediately after removal in a 10% buffered formalin solution for a maximum of 24 h at room temperature before being dehydrated and paraffin-embedded under vacuum. The tissue sections were deparaffinized with EZPrep buffer (Ventana Medical Systems). Antigen retrieval was performed with CC1 buffer (Ventana Medical Systems), and sections were blocked for 30 min with Background Buster solution (Innovex).

The immunohistochemical detection was performed at the Molecular Cytology Core Facility of Memorial Sloan Kettering Cancer Center using a Discovery XT processor (Ventana Medical Systems). All the tissues were harvested from mice and fixed in 4% PFA overnight. Fixed tissues were dehydrated and embedded in paraffin before 5 μm sections were put on slides. The tissue sections were deparaffinized with EZPrep buffer (Ventana Medical Systems), antigen retrieval

was performed with CCI1 buffer (Ventana Medical Systems) and sections were blocked for 30 min with Background Buster solution (Innovex) or 10% normal rabbit serum in PBS (for CAV1 staining). CAV1 sections were incubated with antibodies against caveolin-1 (Cell Signaling, cat. no. 3267, $1\ \mu\text{g ml}^{-1}$) for 5 h, followed by 60 min of incubation with biotinylated rabbit anti-goat IgG (Vector, cat. no. BA-5000) at 1:200 dilution. pMAPK sections were blocked with avidin/biotin block for 12 min, followed by incubation with pMAPK antibodies (Cell Signaling, cat. no. 4370, $1\ \mu\text{g ml}^{-1}$) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Labs, cat.no. PK6101) at 1:200 dilution. Ki67 sections were incubated with Ki67 antibodies (Vector Labs, cat. no. VP-K451, $0.4\ \mu\text{g ml}^{-1}$) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Labs, cat. no. PK6101) at 1:200 dilution. CD31 sections were incubated with CD31 antibodies (Dianova, cat.no. DIA-310, $1\ \mu\text{g ml}^{-1}$) for 5 h, followed by 60 min incubation with biotinylated rabbit anti-rat IgG (Vector Labs, cat. no. PK-4004) at 1:200 dilution. Detection was performed with a DAB detection kit (Ventana Medical Systems) according to manufacturer's instructions, followed by counterstaining with haematoxylin (Ventana Medical Systems), then coverslipped with Permount (Fisher Scientific).

Statistical analysis. Statistical analysis for in vitro and in vivo experiments was performed using GraphPad Prism (GraphPad 7 Software). A two-tailed Student's unpaired t test was conducted to compare control versus treated groups. The significance level was set at $P < 0.05$. Independent experiments were conducted with a minimum of two biological replicates per condition to allow for statistical comparison. Error bars represent the standard error of the mean (s.e.m), and P values are indicated in the figure captions and main text. All cellular experiments were repeated at least two times. Survival plots of experimental metastasis models were analysed using the Mantel–Cox log-rank test. For all in vivo experiments, the sample size was at least $N = 5$ mice per treatment group. These sample sizes were chosen based on previous literature and our own

expertise. All animals were included in analyses. Animal cohorts were randomly selected. Investigators were not blinded. Representative images of tissue histology were from at least five images of the same tissue, and duplicate staining was done for each tissue from experiments so that at least two out of five organs/tissues were embedded and stained.

Life Sciences Reporting Summary. Further information on experimental design and reagents is available in the Life Sciences Reporting Summary.

Code availability. All Python scripts used in the QSNAP analysis, the performance analysis of chemical descriptors and predicted INP formation results of DrugBank small molecule drugs based on the decision tree can be found in the following two repositories, hosted on the collaborative open source software development platform GitHub: 1. nano-drugbank (<https://github.com/choderalab/nano-drugbank>, <https://doi.org/10.5281/zenodo.1117405>). 2. IR783-QSNAP (<https://github.com/choderalab/IR783-QSNAP>, <https://doi.org/10.5281/zenodo.1117393>). A full list of small molecule drugs from DrugBank predicted to self-assemble into indocyanine nanoparticles according to the decision tree analysis is available as a CSV file in the nano-drugbank repository (df_drugbank_decision_tree.csv).

Data availability. The data sets generated and/or analysed during the current study are available from the corresponding authors upon reasonable request.

References

- Nicholls, A. Confidence limits, error bars and method comparison in molecular modeling. Part 1: the calculation of confidence intervals. *J. Comput. Aided Mol. Des.* **28**, 887–918 (2014).
- Schomburg, K. T., Wetzler, L. & Rarey, M. Interactive design of generic chemical patterns. *Drug Discovery Today* **18**, 651–658 (2013).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were chosen based on previous literature in nanomedicine and tumor biology, and our own expertise.

2. Data exclusions

Describe any data exclusions.

No data was excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Myc/b-catenin liver cancer model was repeated 3 times. HCT116 colon cancer xenografts experiments were done 3 times.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

For HCT116 xenografts, we used stratified randomization: 7 days after tumor inoculation the tumor volume was measured and mice were randomized to different cages so that the average tumor volume will be similar in all groups. In the liver cancer model we used simple randomization - meaning the assignment of inoculated mice to a treatment was random.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The survival experiment of liver cancer bearing mice was blinded as the person taking the data did not know which groups were treated or not (done by animal core facility). For tumor volume experiments we didn't use blinding because it's a quantitative objective measurement.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- | | |
|--------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The <u>exact sample size</u> (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement indicating how many times each experiment was replicated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Clearly defined error bars |

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

GraphPad Prism was used to analyze experimental data. All Python scripts used in the QSNAP analysis, performance analysis of chemical descriptors, and predicted INP formation results of DrugBank small molecule drugs based on the decision tree can be found in the following two repositories, hosted on the collaborative open source software development platform GitHub:

1. nano-drugbank (<http://www.github.com/choderalab/nanodrugbank>).
2. IR783-QSNAP (<https://github.com/choderalab/IR783-QSNAP>, DOI: 10.5281/zenodo.1117393)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All materials are generally commercially available without restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

CAV1, R&D Systems, MAB5736, pMAPK antibodies (Cell Signaling, cat# 4370, 1ug/ml), biotinylated goat anti-rabbit IgG (Vector labs, cat#PK6101) at 1:200 dilution. Ki67 (Vector, cat# VP-K451, 0.4ug/ml) , biotinylated goat anti-rabbit IgG (Vector labs, cat#PK6101) at 1:200 dilution. CD31 antibodies (Dianova, cat# DIA-310, 1ug/ml) , biotinylated rabbit anti-rat IgG (Vector labs, cat#PK-4004) at 1:200 dilution.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

All cell lines were purchased from ATCC except for SK136 which was a primary cell line derived from liver tumors from Scott Lowe's Laboratory, Memorial Sloan Kettering Cancer Center.

b. Describe the method of cell line authentication used.

Cell lines were used as purchased directly from ATCC. The MCF-7 cell line is on the ICLAC Register of Misidentified Cell Lines (the contaminating cell line is OVCAR-8), and therefore we used these cells only up to ten passages after receiving the frozen vial directly from ATCC. Our CAV1 immunohistochemistry data for MCF-7 is in agreement with the Broad Institute Cancer Cell Line Encyclopedia mRNA data, but not with OVCAR-8.

c. Report whether the cell lines were tested for mycoplasma contamination.

All cells were tested negative for mycoplasma contamination at the time of experiments.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

The MCF-7 cell line was used due its known CAV1 expression and our extensive previously collected and published data on the permeability/density of 3D tumor spheroids made with this line (Jena, et. al., Carbon, 2016).

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

FVB and athymic nude, nu/nu were purchased from Jackson Laboratories.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Provide all relevant information on human research participants, such as age, gender, genotypic information, past and current diagnosis and treatment categories, etc. OR state that the study did not involve human research participants.