HHS Public Access

Author manuscript

Adv Mater. Author manuscript; available in PMC 2019 October 23.

Published in final edited form as:

Adv Mater. 2019 October ; 31(43): e1902798. doi:10.1002/adma.201902798.

Using Large Datasets to Understand Nanotechnology

Kalina Paunovska[1], **David Loughrey**[1], **Cory D. Sago**[1,2], **Robert Langer**[3,4], **James E. Dahlman**[1]

[1]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory School of Medicine, Atlanta, GA, 30332, USA

[2]Current Address: Guide Therapeutics, Atlanta, GA, 30332, USA

[3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

[4]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

Abstract

Advances in sequencing technologies have made studying biological processes with genomics, transcriptomics, and proteomics commonplace. As a result, this suite of increasingly integrated techniques is well positioned to study drug delivery, a process that is influenced by many biomolecules working in concert. Here we describe how omics-based approaches can be used to study the vast nanomaterial chemical space as well as the biological factors that affect the safety, toxicity, and efficacy of nanotechnologies. We focus on the generation and analysis of large datasets, describing methods to interpret them. We also describe how these datasets have been applied to nanomaterials to date. Finally, we propose new ways sequencing datasets can answer fundamental questions in nanotechnology-based drug delivery.

A2. Drug delivery is a complex process involving many biomolecules

Biological processes are carefully regulated. For example, proliferation is not governed by a single master gene. Instead, it is influenced by post-translational modifications, transcription factor binding sites, RNAs, proteins, lipids, carbohydrates, and combinations thereof[1]. The same is true for cell death[2], metabolism[3], and endocytosis[4]. This biological complexity is critical to cell function. However, complexity makes it difficult to deconvolute how individual biomolecules contribute to a phenotype. The scale of biological systems makes this problem more difficult. As an example, the human genome consists of approximately 20,000 protein coding genes that interact with one another dynamically and in response to environmental cues. Even if we ignore the approximate 98% of the human genome that does not encode protein-coding genes[5], the complexity of the genome is a universal problem for the biomedical field[6]. Nanomedicine delivery is a complex process regulated by the body[7]. Successful *in vivo* drug delivery requires a nanoparticle to protect the drug from

Correspondence: james.dahlman@bme.gatech.edu.

Conflict of Interest Statement. J.E.D. and C.D.S. are co-founders of Guide Therapeutics.

degradation, avoid the systemic immune system, avoid clearance organs, enter the desired tissue, select the right cell type within a complex tissue microenvironment and - if the drug requires cytoplasmic delivery - gain access to the cytoplasm without degrading in an organelle (Figure 1A). At each step the nanomedicine must overcome defenses that have evolved to sequester and degrade foreign materials; this makes drug delivery inefficient. For example, a lipid nanoparticle (LNP) that delivers small interfering RNA (siRNA) to hepatocytes in mice, non-human primates, and humans[8] was used to ask an important question: if a LNP carrying siRNA reaches the endosome of a target cell *in vivo*, what percentage of the siRNA accesses the cytoplasm? This LNP only released 2% of its siRNA into the cytoplasm[9]. Recognizing these inefficiencies, clinical advances in nanotechnology research[8, 10–13] are impressive. However, despite these advances, leaders have called for changes to the way nanotechnologies are studied or described[14–16]. Our experience supports these calls for change.

Nanomedicines have untapped potential, in large part because they are still difficult to design a priori, and like all drugs[17], are affected by biological interactions that are hard to study. However, developments in next-generation sequencing technologies (NGS) are allowing biologists to answer questions on an entirely new scale (Figure 1B). Although the definition of 'big data' varies[18], the ability to generate and analyze large biomedical datasets could help study fundamental nanotechnology questions. Namely, how does nanoscale chemical structure influence drug delivery *in vivo*? And, which biological pathways govern nanoparticle delivery *in vivo*?

Over 40 years of work has resulted in a substantial body of knowledge[19] describing interactions between nanotechnology and biomolecules. For example, evidence shows that the high surface area to volume ratio of nanoparticles makes it thermodynamically likely[20] that diverse molecules will bind nanoparticles after they are administered[21]. The composition of this 'corona' changes with time[22] and local environment[20, 23]. These interactions alter how nanoparticles engage the immune system or target cell[24]. In one example, a LNP was bound by serum apolipoprotein E (ApoE), which increased delivery to hepatocytes, which were the target cell type[25]. In another, the protein corona blocked interactions between transferrin-targeted nanoparticles and receptors on the cell surface, thereby reducing delivery[24]. We also know that physical barriers influence nanoparticle delivery. Cationic nanoparticles can be disassembled by the renal anionic basement membrane[26], and nanoparticles inefficiently access healthy brain parenchyma due to the blood brain barrier[27]. By contrast, nanoparticles access hepatocytes easily due to porous endothelial cells, discontinuous basement membranes in hepatic sinusoids[27], and slowed blood flow that increases nanoparticle extravasation[28].

We also understand that specific genes can affect nanoparticle delivery (Figure 2). Most studies to date have identified genes that alter nanoparticle or nucleic acid endocytosis *in vitro*[9, 29–32]. In a recent example, authors manipulated cells with small molecules that manipulated genes, then administered LNPs carrying mRNA. The authors found that small molecule drugs altered mRNA delivery; some drugs improved mRNA delivery, whereas others reduced it[33]. Publications have also studied how genes alter nanoparticle delivery *in vivo*. These results suggest that specific genes can alter systemic nanoparticle

pharmacokinetics, biodistribution[34], and endocytosis[25]. For example, Bertrand et al. quantified how nanoparticles with high (or low) amounts of poly(ethylene glycol) (PEG) circulated in genetic knockout mice. They found that the low-density lipoprotein receptor (LDLR) played a dominant role in nanoparticle clearance, irrespective of PEG content[35]. In a second example, it was found that Caveolin 1, a gene that is critical for caveolin-mediated endocytosis, was needed for LNPs to enter endothelial cells and liver macrophages, but was not important for delivery to hepatocytes, or macrophages in other tissues[36]. These results suggest that inhibiting caveolin signaling may retarget nanoparticles *in vivo*. Separate studies have found that genes related to mRNA translation[33], lysosome formation and maturation[37], and anti-viral immune response[38] can also alter nanoparticle delivery. Finally, systemic physiology can alter delivery. For example, delivery to non-tumor organs varied when nanoparticles were administered to healthy and tumor-bearing mice[39]. Similarly, the administration of the anti-malarial drug Chloroquine reduced nanoparticle uptake by macrophages[40], and nanoparticles delivering rapamycin can increase the tolerability of biologics[41] in mice and non-human primates.

Finally, it is accepted that the interactions between nanomedicines and different molecules vary with nanoparticle chemical composition[42], shape[43], and size[44]. Given that specific genes, systemic physiology, and nanomedicine chemical structure come together to dictate nanomedicine behavior, many interesting questions remain unanswered. For example, whether there are master regulatory genes that affect many types of nanoparticles; if there is a ‘p53 for nanoparticle delivery’, it has not been identified. Luckily, complexity is a biological norm, and new sequencing technologies are well positioned to help us study interactions between nanomaterials and the body.

A3. Studying the nano-bio interface using next generation sequencing

Next generation sequencing approaches enable single cell and multiomic analyses

A suite of technologies based on high-throughput NGS have been created and validated. All of these are driven by advances in sequencing-by-synthesis, which allows scientists to characterize millions of molecules at the same time. These omics techniques, referred to as “sequence census” methods, can examine the genome (DNA), transcriptome (RNA), and epigenome (DNA modifications). All exploit the fact that DNA sequences can function as a digital substrate that is easily counted[45].

These technologies have evolved rapidly. Soon after NGS was reported, scientists designed ways to sequence DNA[46], and later, RNA[47], from single cells. Advances in single cell required specific advances in acquiring and analyzing data. In particular, when acquiring single cell RNA-seq (scRNA-seq) data, it is important to understand ‘drop-out’, an effect wherein datasets contain many genes with no expression. By developing standardized methodologies to overcome drop-out, single cell techniques have enabled targeted RNA[48–49] and whole transcriptome[50–51] analysis. By sequencing RNA from single cells, scientists improved their fundamental understanding in many fields of biology, examining everything from the diversity of microbial ecosystems to the intratumor heterogeneity and clonal evolution of human cancer[52–53]. As an example, scRNA-seq studies have been used to differentiate subclasses of a given cell type (e.g. neurons[54–56], or immune cells[57–59]), or

study heterogeneous cell responses to a given biological stimulus[60]. In one representative example, Villani *et al.* performed unbiased scRNA-seq on 2400 peripheral blood mononuclear cells. By analyzing the subsequent gene expression data, they identified new subtypes of dendritic cells and monocytes in human blood, enabling more accurate immune monitoring in disease[57]. In order to generate these single cell data, authors combined an experimental and computational strategy to identify discriminative surface markers in clusters of cells that were similar to each other. They isolated the cells using these markers and validated the identity of these inferred subtypes using scRNA-seq. In order to ensure the data were robust, the authors corroborated their findings by analyzing peripheral blood mononuclear cells from ten independent healthy individuals. Although scRNA-seq approaches are not frequently used to study cellular response to nanomaterials, we are optimistic this approach will be important to the nanomedicine field for 2 reasons. First, scRNA-seq is now easy to use. In fact, there is an ongoing effort called the 'Human Cell Atlas' that aims to perform scRNA-sequencing on as many cell types as possible[61]. Second, in the papers cited above, authors found that a collection of cells thought to be homogenous exhibit a high degree of genetic and functional heterogeneity. These data suggest that gene expression and subsequent cell function, even within a given cell type, exists on a spectrum. These approaches could similarly reveal subtypes of immune cells that readily interact with nanomaterials. By studying the different gene expression profiles in immune cells that do (or do not) respond to nanotechnology, master regulatory genes that trigger immune responses to nanomaterials or promote effective endosomal release may be identified.

More recently, the integration of diverse platforms (multiomics) has begun. In these examples, large scale analysis of multiple biomolecules is performed[62–64]. One key aspect of multiomic data generation is the fact that scientists must (i) process cells and (ii) design sequencing pipelines that allows several datasets to be acquired. In one example, scientists measured the genomic copy-number variations, transcriptome and DNA methylome of 25 single cancer cells. The authors were able to acquire these multiomic data using a gentle lysis procedure that dismantled the cellular membrane of an individual cell while keeping the nucleus intact. This preserved nucleus was used as a substrate for single cell DNA methylomic analysis, while the cytoplasmic lysate was used to acquire transcriptomic information from the same cell. They identified two distinct subpopulations within these cells and showed the transcriptomic heterogeneity within each subpopulation[65] affected cell function. In another example, scientists used NGS to concurrently measure transcriptomic and epigenomic data, in order to evaluate the mechanisms of neurodegeneration in Alzheimer's disease, and how the environment and the genome act through different cell types[66]. Once again, the authors used a novel experimental approach to acquire the data; more specifically, the authors performed in parallel chromatin immunoprecipitation and RNA sequencing on harvested mouse hippocampus. This allowed seven different epigenetic modifications that mark distinct functional chromatin states and the corresponding changes in gene expression to be analyzed simultaneously. By profiling transcriptional and chromatin state dynamics, they found that immune-cell-specific enhancer regions and response genes were more accessible to transcription factors, suggesting the pathogenic capacity of the immune system in Alzheimer's disease. A coordinated decrease in synaptic plasticity genes

was also found, linking these multiomic readouts to a potential mechanism of disease progression.

The coupling of protein mass spectrometry to genomics, known as proteogenomics[67–68], is another new class of technologies to generate multiomic datasets. Although mass spectrometry has analytical limitations[69], these are being addressed. To date, proteogenomics has been applied to traditional biological problems. For example, scientists characterized human colon and rectal cancer[67]; using proteogenomics, the authors identified 4 subtypes of diffuse gastric cancers, associated with proliferation, immune response, metabolism, and invasion, respectively[70]. However, through these studies, best practices have been established that provide a framework to characterize protein-nanomaterial interactions. Thus, proteogenomics has the potential to be applied to the protein corona and other interactions between nanomaterials and proteins.

Although multiomics approaches have not – to date – been applied to nanomaterials, these techniques permit scientists to characterize complex cellular responses[71–72]. It is therefore very likely that multiomics can help elucidate how cells respond to nanomaterials.

Transcriptomics can uncover how cells respond to nanoparticles

In contrast with multiomics, transcriptomics has already been used to interpret the complex effects that nanomaterials and biomaterials have on gene expression. There are a number of recent examples of the nanotechnology field taking advantage of transcriptomics, both *in vitro* and *in vivo*. Carrow *et al.* recently used RNA-seq to identify more than 4000 genes whose RNA expression changed when human mesenchymal stem cells (hMSCs) interacted with nanosilicates[73]. Notably, they found that particular signaling pathways were upregulated, including the stress-responsive and surface receptor-mediated mitogen-activated protein kinase (MAPK) pathways. The authors also characterized a number of biophysical and biochemical cellular behavior and found that nanosilicates promote stem cell osteochondral differentiation. In particular, by analyzing changes in genes that are part of biological pathways related to osteogenesis, researchers saw that hMSCs exposed to nanosilicates tended to favor bone and cartilage lineages. They found that genes such as cartilage oligomeric matrix protein, aggrecan, and collagen type I α1 chain were up-regulated; these genes are characteristic of osteochondral differentiation. Taken together, these data suggest that proliferation and differentiation pathways were influenced by nanomaterials. As another example, Feliu *et al.* utilized primary human bronchial epithelial cells to show that cationic dendrimers caused significant changes in gene expression, even at doses that did not lead to acute or overt signs of cytotoxicity[74]. After administering a dose of 0.1 μM PAMAM dendrimers – which translates to a dose of roughly 1.4 μg / mL *in vitro* – to these cells, they found that the expression of 203 genes changed. Interestingly, by performing gene ontology enrichment analysis, the authors found that many of these genes were part of pathways related to cell division and cell cycle regulation. The authors created network diagrams to visualize predicted impacts on downstream pathways after upregulation and downregulation of specific genes. These results are important, given that many studies rely on overt assays to screen for nanoparticle toxicity. The results may also have implications for tumor-targeted nanoparticles, since tumor growth can be driven by aberrant

cell division and cell cycle regulation. In another example, Lucafò *et al.* reported the interaction of fullerenes with human MCF7 tumor cells showing that they cause a time-dependent alteration of gene expression, arresting cell cycle progression and promoting the entry in G0 phase[75]. By performing whole-transcriptome RNA-seq analysis on cells exposed to fullerenes, the authors found that mTOR signaling, which regulates cell growth and proliferation, was inhibited while genes upstream of TGF-β, important for cell remodeling and adhesion, were upregulated – suggesting that nanoparticles can alter cell cycle regulation. In addition, Gliga *et al.* showed that cerium oxide nanoparticles negatively affect neuronal differentiation and interfere with cytoskeletal organization in the murine cell line C17.2, which can be used as a model for developmental neural stem cells[76]. Cerium oxide nanoparticles were known to show cytoprotective effects. However, by analyzing gene expression using RNA sequencing this study found that the expression of at least 795 genes changed over a 7 day period after C17.2 cells were exposed to nanoceria. Changes in gene expression were compared to changes elicited with a common anti-oxidant, N-acetylcysteine, and samarium-doped nanoceria, which has previously been shown to have lower antioxidant activity than nanoceria alone. Notably, the authors found that nanoceria inhibited neuronal stem cell differentiation extensively, compared to N-acetylcysteine and samarium-doped nanoceria, when they analyzed the genes that were changed, illustrating that antioxidant properties were not necessarily beneficial in all cases. In *Chlamydomonas reinhardtii*, a model organism, authors found that exposure to four different commonly used metal nanoparticles – nano-Ag, nano-TiO_2, nano-ZnO, and CdTe/CdS quantum dots (QD) – had both similar and relatively distinct effects on the transcriptome. More specifically, Zn, QD and Ti based nanoparticles had upregulation and / or downregulation of similar genes, whereas Ag elicited an opposite transcriptional response in *Chlamydomonas reinhardtii* when compared to the other three nanoparticles. Notably, some of the changes included potential proteasome inhibition which could suggest interest as a cancer chemotherapy agent[77]. Also in *C. reinhardtii*, Beauvais-Flück *et al.* showed that up to 4784 transcripts were dysregulated when exposed to subnanomolar methyl-mercury even after two hours. Genes involved in cell motility, nutrition, and photosynthesis were among the main regulated transcripts highlighting the tolerance mechanisms for microalgae at sublethal methyl-mercury concentrations[78]. Finally, additional evidence that nanoparticles alter genome-wide gene expression has been found *in vivo*; engineered iron sulfide nanoparticles were shown to cause substantial gene expression alterations in pathways related to immune and inflammatory responses, detoxification, oxidative stress and DNA repair and damage, in adult zebrafish[79]. These results illustrate that major transcriptional changes can be tracked *in vivo* when an organism is exposed to a nanoparticle. These examples are complemented by evidence suggesting the composition, size, or shape of a biomaterial potentiates the cellular response to that material[80]. Studies that record the cellular response to biomaterials have been collated in the Compendium for Biomaterial Transcriptomics (cBiT)[81], a collection of transcriptional profiles of cells after biomaterial exposure; this resource will likely continue to become even more valuable as more data become available.

As demonstrated by the studies above, best practices for RNA-seq data generation and gene expression analysis are established[82]. The first step is to clearly define a biological question. One simple test case would be ‘What RNAs are affected by a given nanomaterial, and can

the RNAs identify a specific cellular signaling cascade that responds to that nanomaterial?'. Second, extract the cellular RNA and convert it to a countable pool of complementary DNA (cDNA) via reverse transcription using polydT or random hexamers using standard kits. Third, sequence this pool of DNA using NGS. Fourth, perform quality control analyses on the data in order to statistically correct biases that arise during sample preparation or sequencing. Fifth, analyze the 'clean' data using an appropriate bioinformatics pipeline, thereby identifying genes with up- or down-regulated expression in response to the nanomaterial[82]. Sixth, use network analysis or cell ontology based approaches to understand whether alterations in gene expression can identify cellular pathways altered by the nanomaterial. Finally, once pathways are identified, it is feasible to make predictions about how the nanomaterial will affect the cellular phenotype (cell growth, death, toxicity, etc.).

A4 Methods to analyze large datasets appropriately

As the output from sequencing platforms reaches the order of terabytes (and billions of sequencing reads), it will be increasingly important to visualize and interpret the data related to biomaterials using best practices. Here we describe common issues faced when interpreting data sets of this size, as well as ways to ensure the data interpretation is appropriate[82–83]. One important consideration when analyzing large data sets is dimensionality. For example, some transcriptomic studies can have 20,000 dimensions; each dimension is the expression of a gene. Given that visualizing data on 20,000 axes is not feasible, datasets are reduced to a smaller number of dimensions so they can be visualized (Figure 3A). High-dimensional objects are replotted in a low-dimensional map; individual objects are represented by a point, and objects that behave similarly are 'clustered' nearby. In addition to making data easier to interpret visually, reducing dimensionality can be used to identify important variables in a complex, multivariable experiment.

PCA allows dimensionality reduction

Dimensionality reduction is often performed using principle component analysis (PCA)[84–85]. Put succinctly, PCA provides a statistical framework whereby the maximum amount of variance is captured with the lowest possible number of dimensions. In biological experiments, where there are usually many more observations than variables, the number of principle components (PCs) is the same as the number of variables. The PCs are sorted by their statistical importance. For example, suppose factors contributing to the cost of a car were studied by generating a dataset with the cost, size, brand, color, and number of wheels of different vehicles. Since all cars have 4 wheels, this variable will not contribute to the variance in car costs. However, the cost might matter, as might the brand, and these two factors co-vary. In this case, principal component 1 (PC1) would be the linear combination of variables that contributed the most amount to variance (e.g. PC1 = 4*cost + 2.4*brand + 1.1*size + 0.3*color + 0.001*num. wheels). In this linear combination, the number of wheels negligibly contributes to the variance, and is therefore unimportant. Then, after factoring in PC1, a second set of relationships can be seen, where (for example) the size and color might co-vary: (e.g. PC2 = 2*size + 1.2*color + 0.7*brand + 0.1*cost + 0.001*num. wheels). Every factor contributes to each PC, but only the factors that explain a lot of the variance and are correlated have high weights for the same PC. In the case of studying

nanomaterial-biological interactions, the factors may be the sets of genes that are up or down-regulated in response to a specific nanomaterial. One important limitation to PCA however is that relationships between variables are often non-linear. In addition, PCA is usually specific for each dataset, making it difficult to compare PCs across studies. As a result, when considering whether a nanomaterial dataset can be analyzed with PCA, it is best to consult an expert in data analysis.

Even with these nuances, PCA can still be used effectively to reduce dimensionality. In biological applications, applying PCA to data with N variables will generate N PCs; if the first PC is responsible for a large percentage of the variance, the dimensionality of the dataset can be reduced by excluding PCs with much smaller contributions. PCA is commonly applied to biological datasets in order to identify experimental conditions that drive variance in gene expression[85]; in a typical gene expression profiling experiment, the first 5 PCs drive up to 50% of the variance, while the remainder explain just one or two percent of the variance and can be ignored. As a result, although nuances in the data can be lost during dimensionality reduction, the general structure of the dataset is preserved. As PCA is applied to nanomaterials, experiments will need to be designed in order to maximize the number of repetitions per observation. Another useful feature of PCA is that once the PC is identified, it can help identify what drives similarities among samples, and remove unimportant sources of variation. Supervised and algorithmic options for analyzing these factors are widely used in transcriptomics[86–87], and therefore, should be applied to nanomaterial datasets.

Applicability of PCA to biological datasets

Currently, PCA is used in biology to answer questions related to (i) genetic differences between cell populations or (ii) gene importance when it comes to understanding a cellular response to specific stimuli. This can be closely related to nano-bio interactions, which would replace a normal biological stimulus (e.g., a cancer drug) with a nanomaterial, thus allowing scientists to probe mechanisms behind these interactions. However, since PCA is easy to perform, it can be applied to datasets inappropriately[88]. For instance, PCA is typically not useful when (i) the variance is somewhat evenly distributed among the principle components, and (ii) the dataset is small and the amount of variables and variance within the dataset is large. What constitutes an appropriately large nanomaterial dataset? As larger datasets are generated using nanotechnology, this question will need to be addressed. Once again, consulting with scientists who specialize in PCA will be important for nanomaterial labs. However, lessons from biological studies may help answer the question. It is generally accepted that biological studies with a large number of replicates can be analyzed with PCA, whereas studies with a small number of biological replicates (e.g., N=3 or fewer), and therefore, relatively high experimental variability, cannot. As a control, biological replicates should cluster together. The larger the number of variables being analyzed, the more technical and biological replicates are required to make statistically powered statements about data. For biological, and nano-related applications, biological replicates should be strongly correlated. Minimizing biological variance within an experiment is also crucial to correct analysis of data. For example, when analyzing nanoparticle delivery data, it will be necessary to separate cells that had ‘low’, ‘medium’,

and 'high' levels of delivery, in order to obtain interpretable data. Given that the absolute values of low, medium, and high can vary with the type of drug being delivered, nanotechnologists will need to provide the rationale for their selection clearly. The advantages and limitations of PCA, as well as best practices, have been reviewed in other fields[84–85]. These best practices will be a useful starting point for nanotechnologists.

Alternative forms of dimensionality reduction

PCA is a dimensionality reduction technique that is mathematically designed to identify axes with maximum variance. However, in some cases, preserving small differences between similar objects is preferred[89]. For example, single-cell sequencing experiments regularly reveal heterogeneity amongst cells that were previously thought to be homogenous[90–91], and often identify important rare cell subpopulations. For example, Shalek *et al.* found that the core antiviral response in pathogenically stimulated primary mouse bone-marrow-derived dendritic cells was coordinated by only a small proportion of the population[92]. In particular, the group found that only 0.8% of the 1700 sequenced cells exhibited antiviral gene expression very early, thereby leading to a larger response from the entire population. Given that immune cell subpopulations have been found in many other biological contexts, these approaches may be useful in overcoming three key limitations to nanomaterials. First, nanomaterials are cleared by circulating immune cells as well as immune cells within tissues. We find it likely that subsets of immune cells – driven by particular signaling pathways - respond more 'aggressively' to nanomaterials. Understanding these pathways could lead to pre-emptive, transient interventions designed to reduce nanoparticle toxicity. Second, nanoparticles can interact with cells via surface receptors. It is feasible that cell subpopulations express higher levels of a given surface receptor, thereby making it easier to specifically target that cell subtype. Third, since many nanoparticles enter cells via endocytic pathways, escaping the endosome is critical. It may be possible to identify cell subsets that are particularly amenable to drug delivery, simply due to the expression of genes related to endosomal escape. In order to identify cell subpopulations with these phenotypes, the best practice would be to analyze single cells, measuring immunostimulation, biodistribution, or cytoplasmic release, and, at the same time, measuring the transcriptomic profile of the cell. In these experiments, it would be important to group cells so small differences between cell types are preserved. For such situations, algorithms like t-distributed stochastic neighbor embedding (t-SNE) are appropriate. T-SNE, first described by Maaten and Hinton in 2008[89], has allowed researchers to analyze cell heterogeneity in new ways[90–91]. Algorithms to visualize t-SNE plots have been adapted for use in multiple languages, including R, python, and MATLAB, making the technique easy to use. Biological predictions made by t-SNE have also been validated using traditional biochemical techniques. For example, DroNc-seq, a method that combines single cell and single nuclei RNA sequencing, was used to identify distinct cell populations with t-SNE. These populations were then confirmed using immunohistochemistry and other methods[90]. t-SNE is useful as an alternative cell clustering and visualization tool when trying to understand cell response to nanomaterials.

Although t-SNE has generated validated predictions when used correctly, it can also be used to draw incorrect conclusions. t-SNE plots are generated using several input parameters, most notably perplexity and the number of iterations run[93]. Authors have shown that

selecting incorrect input variables can lead to images that contain clusters when in fact no clusters exist[93] (Figure 3B); these are analogous to false positives. Moreover, every time a t-SNE plot is generated, the plot changes slightly, since all t-SNE plots are stochastic[89]. As a result, although the general structure is preserved and has meaning, interpreting relationships between individual points on the plot is inappropriate since the position of each individual point varies each time the analysis is performed (Figure 3C).

Analyzing biological datasets with unbiased clustering

A second approach used to analyze large datasets is unbiased clustering. Unbiased clustering helps visualize experimental groups that performed more similarly to one another than they did to other groups, without losing any information. Since clustering algorithms rely on different mathematical assumptions, it is important that clustering is performed with the appropriate algorithm, and that altering the algorithm does not dramatically alter the clustering pattern[83]. The most common algorithms are hierarchical, centroid/partition (e.g. k-means), density-based (e.g. DBSCAN)[94], and self-organizing maps (SOMs)[95]. In k-means clustering, the user selects a k value based on the number of clusters that the data will be partitioned into. If the user expects there to be many clusters, a high k number is selected; if the user expects few clusters, a low k number is selected. The algorithm associates nearby values based on their means; as more values are associated, the mean of all the values becomes the new mean until k clusters are formed[96]. Conversely, DBSCAN clusters are based on how closely points pack together and outliers are determined based on their presence in low density regions[94]. When measuring how cellular mRNA expression changes with response to a drug (or a biomaterial), hierarchical clustering or SOMs are often used. The appropriateness of a given clustering algorithm depends on the size and complexity of the dataset, as well as the research question being asked[97–98], and guides to select the correct clustering algorithm have been published[99–100]. Using appropriate clustering algorithms when analyzing biomaterial data will be important. For example, if k-mean clustering is employed, how is the number of clusters selected? Scientists studying biomaterials can learn from examples in other fields[101]. Unbiased clustering has been utilized in order to analyze how cells cluster based on nanoparticle functional delivery as well as how nanoparticles cluster based on material properties[34, 102]. Given enough of this type of data, these analyses could be instrumental for intuitively designing future generations of nanoparticles.

To help evaluate whether the data are suitable for a given clustering algorithm, validation algorithms have been developed. Validation algorithms are based on metrics that evaluate how tight data within a given cluster are, and what the distance between clusters is[103–104]. Validation algorithms are often subdivided by the type of clustering they employ; these include compactness, separation, and connectedness[104]. For example, to validate k-nearest neighbor clustering, a validation algorithm was developed based on the following idea: if we take a data point from a cluster, its k-nearest neighbors should be in the same cluster[103]. Put simply, the k-nearest neighbor is determined by assigning a value to each object; the value is proportional to its distance from the object. Then, depending on the k constant, the objects are group based on closeness; when k = 1, the nearest neighbor is clustered with the object of interest.

A5. Visualizing large datasets

Network diagrams for visualizing complex interactions

Additional techniques are then required to visualize large datasets. Two common methods of data visualization are network diagrams and heatmaps. Network diagrams integrate data from many sources to model interactions within a biological system. As an example, scientists generate networks combining gene expression and other omics data[105]. Since looking at raw network diagrams can be challenging, they are simplified using algorithms that cluster the raw network[105]. This clustering utilizes gene expression data to quantify correlation values between genes. If the expression of A and B always change in the same direction, the algorithm tends to cluster them together. Given that even these clustered networks can be difficult to interpret, manual editing of the network diagram can be employed to emphasize a specific component of the biological pathway. Alternatively, the gene expression may be overlaid on validated pathways using the Kyoto encyclopedia of genes and genomes (KEGG)[106–108] or the gene ontology consortium[109–110]. These network diagrams – which are visual and qualitative – are also often augmented by including quantitative metrics derived from the dataset. As an example, information from gene or protein expression profiles can be included in network diagrams by making over or under-expressed genes/proteins stand out on the network. A common tool for creating integrated network diagrams is Cytoscape[111].

One related question that will need to be addressed as network analyses are used to understand biomaterial / cell interactions is the extent to which subtle biological interactions matter. In some cases, studying single genes will suffice. For example, the gene ApoE was shown to dramatically impact the delivery of a lipid nanoparticle *in vivo*; with ApoE, the nanoparticle was effective, and without it, the nanoparticle stopped functioning entirely[25]. However, it is likely that most nanoparticle-biological interactions will be driven by many genes interacting with one another. In the cases where many genes influence delivery, network analysis could focus on interactions between genes involved in endocytosis, metabolism, or intra / intercellular transport. To understand how many genes work in concert, network diagrams can be used to show interactions between hundreds or thousands of genes in a more unbiased way. Once these interactions are identified, scientists can evaluate whether the individual interactions are synergistic, additive, or antagonistic. If two genes interact synergistically, their effect on a phenotype is greater than the sum of each gene's individual impact. If they interact antagonistically, their effect is less than if they were additive. Importantly, it is possible to evaluate how single genes and collections of genes can synergize or antagonize one another in a biological pathway[112].

Using heatmaps to highlight differences within a dataset

Like network diagrams, heatmaps can be used to qualitatively highlight regions of interest in multivariate data. For example, gene expression heatmaps can identify genes that have high and low expression profiles if they cluster. If a clear and broad pattern exists within a dataset, heatmaps can highlight that pattern. Heatmaps are regularly used to supplement biological analyses. As an example, Subramanian *et al.* used hierarchical clustering to compare how 6 human cancer cell types clustered when analyzed using their profiling

method, L1000, compared to Affymetrix and Illumina microarrays, and NGS-based RNA-seq, showing that each cell type clustered with itself independent of the sequencing/profiling system used[113]. They also analyzed 3333 drugs and 2418 additional compounds and showed that many of the drugs had potential off-target effects and potentially acted on multiple pathways. Honing in on the histone deacetylase (HDAC) superfamily of proteins, they were able to cluster inhibitors based on their selectivity for 13 different HDAC proteins[113]. Similarly, Hughes *et al.* assessed the effects of 300 different mutations and chemical treatments on *S. cerevisiae* and used hierarchical clustering to show that subtle changes in expression profiles can be tolerated and studied[114]. This is especially useful when looking at the effects of knocking out uncharacterized genes on a variety of cell processes. Heatmap analysis of sequencing data can be useful for identifying how a gene's expression changes over time in response to a biomaterial, and has been used to identify nanoparticles that efficiently deliver drugs[102, 115], identify cell types that are targeted by similar nanoparticles[102], and to identify nanoparticle chemical properties that tend to promote *in vivo* delivery.

Best practices for data visualization tools

Like other big data tools, it is important to ensure heatmaps are interpreted correctly. As an example, heatmaps use color to denote differences between samples; but the same color looks different when placed next to different colors[116] (Figure 4a). In addition, data can be scaled by row or by column – this decision is dictated by what differences are being emphasized within a dataset. For example, a test dataset may have 'cell types' as column labels and 'genes' as row labels. The scaling method will dictate whether differences in the expression of one gene throughout multiple cell types (scaling by row), or differences in multiple genes' expression throughout one cell type (scaling by column) is emphasized. Attempting to qualitatively interpret data between rows if scaling colors by row or between columns when scaling colors by column would be incorrect – the colors may appear similar, but the absolute values would differ (Figure 4B). Similarly, if the dataset has many more dimensions in one variable (e.g., genes) than another variable (e.g., cell types), it is best to cluster by the variable with fewer dimensions[83]. For example, if the expression of 20,000 genes is analyzed in 80 cell types, it is better to cluster by cell type first. Finally, data normalization (e.g. centering/scaling data around the mean, median, standard deviation (STD)) as well as the method used for clustering (e.g. Ward's, average, single, or complete) can change how the data cluster (Figure 4C). Finally, it is important to avoid dropping samples from the dataset, since this can have a large effect on how the rest of the samples cluster, as well as how the data is normalized. By understanding the limitations of over interpreting the color of a single box, running the data through more than one clustering algorithm (to ensure the clustering pattern does not dramatically change), analyzing the colors within the right 'direction' (i.e., column or row), and avoiding dropping data from the dataset, heatmaps can be generated that provide compelling evidence of trends within complex biological systems; in many cases, these trends would be difficult to identify using other methods.

It is similarly important to understand the variance associated with your large data set; variance can be biological or technical. Biological variance is understood and can largely be

mitigated by using a large number of replicates. Technical variance is still less well understood and can change with the experiment. As an example, reverse transcribing RNA can lead to bias that alters RNA-seq datasets[117]. Scientists also found that specific sequencing machines perform differently[118] and can generate bias[119]. There are simple ways to minimize variance. For example, including a sufficient number of biological replicates, and including appropriate positive and negative controls. One additional control that is important to consider when analyzing many biomolecules at once is the 'input'. For example, if you administer a pool of DNA-barcoded cells to an animal, it is important to sequence that 'input' pool, so you can normalize your output appropriately. Finally, any hits identified with any initial high-throughput screen should be independently validated using a tool like quantitative PCR, although previous studies have shown high correlation between the two techniques[51, 120–121].

A6. Future perspectives

High-throughput data generation and analysis is not without difficulties, but this does not downplay its potential impact on nanomedicine. Recent clinical results using nanomedicine are cause for great excitement; these advances can be furthered using sequencing technologies. For example, nanoparticles carrying small molecules have been safely administered to patients[12], and siRNA delivered to hepatocytes by GalNAc conjugates[13] or lipid nanoparticles[8] have treated genetic disease. At the same time, the need for systemically administered nanomedicines that target non-hepatocytes is significant, since most systemically administered drug delivery systems are still sequestered in the liver. The need for drug delivery is also growing. Traditional small molecule therapies have been joined by drugs based on proteins, siRNA, miRNA, mRNA, lncRNA, ASOs, ZFNs, TALENs, and CRISPR-Cas proteins. Each class of drugs will present numerous opportunities for nanotechnologists; as an example, the nanoparticle formulation that delivers a Cas9 mRNA is unlikely to be the best nanoparticle formulation for a Cas9 ribonucleoprotein. One additional example is whether the design rules for nanomedicines delivering one drug class (e.g., small molecules or proteins) will pertain to nanomedicines delivering another drug class (e.g., siRNA or mRNA). On one hand, it is possible to foresee a gene acting as a semi-master regulator of drug delivery. On the other, the biological response to nanoparticles containing proteins may be entirely different than the biological response to nanoparticles containing nucleic acids.

Using NGS, scientists can now quantify how thousands of nanoparticles target cells directly *in vivo* by formulating nanoparticle to carry rationally designed 'DNA barcodes'[34, 36, 102, 122–125] (Fig. 5A). In a separate example, scientists have used non-NGS forms of DNA analysis to perform high throughput *in vivo* assays of chemotherapy delivery[126] (Fig. 5B). These high throughput *in vivo* studies may eventually relate nanomaterial structure to *in vivo* delivery. However, future advances still need to be made, particularly in the ability to perform multivariate analysis on these large datasets. For example, when one of the components making up the nanoparticle is varied (e.g., poly(ethylene glycol), PEG), interpreting causality in the dataset is difficult. If two nanoparticles with varying PEG molar ratios are tested, and the nanoparticle with high molar percentages of PEG performs well, is it due to increased PEG, or decreased non-PEG

components? PCA, t-SNE and other dimensionality reduction techniques are equipped for complex analyses like this. If this high throughput *in vivo* approach is coupled to an improved mathematical framework that permits scientists to understand how multivariate changes in nanoparticle structure alter delivery, nanoparticles with improved traits can be designed. For example, one key limitation in nanoparticle delivery is the unwanted clearance by immune cells, particularly in the liver and spleen. By quantifying how thousands of chemically distinct nanoparticles deliver drugs to on-target cells as well as these off-target cells, scientists may be able to 'evolve' nanoparticles that interact with clearance organs less frequently.

One way sequencing may improve nanomedicine is by making the pre-clinical 'pipeline' used to discover nanoparticles more efficient. For example, the standard in the field is to synthesize chemically distinct nanoparticles, screen them *in vitro*, and select a small number of compounds for *in vivo* studies. However, *in vitro* nanoparticle delivery can be a poor predictor of systemic *in vivo* nanoparticle delivery[102]. At the same time, certain *in vitro* systems that recapitulate organ physiology may predict *in vivo* delivery. We envision high throughput studies comparing *in vivo* delivery to organ-on-chip systems[127] using thousands of nanoparticles[124]. By statistically comparing how thousands of different nanoparticles behave, these studies could elucidate the engineering (or biological) variables that make organ-on-chip systems predictive of *in vivo* behavior. A second inefficiency in the nanoparticle discovery pipeline is the unknown relationship between nanoparticle delivery in a mouse, and nanoparticle delivery in a rat, pig, non-human primate, or human. A systematic study of small animal models designed to identify a 'gold standard' animal to predict delivery in large animals has not been reported; this would constitute a significant advance for the field. We anticipate these studies may reveal that a given nanomedicine behaves differently in different mouse strains. Mouse strain-specific delivery has been observed with a promising virus[128–129] selected using a novel *in vivo* viral evolution based approach[129]. The correct pre-clinical animal model may also change with the desired tissue; as an example, compared to mice, ferrets are better models for human airborne viral transduction[130]. By testing thousands of nanomaterials *in vivo* and understanding how strain- and species-dependent biological factors influence delivery, these large datasets may help improve how well pre-clinical models predict delivery in humans.

Big datasets may also be useful for understanding how to design nanotechnologies. For example, a method for de novo protein design[131] was recently reported; using machine learning, Butterfield *et al.* created a large library of protein-based nanocages (Fig. 5C). By applying selection pressures, nanocages were evolved using a 'bottom-up' approach to carry their own mRNA genome. Specifically, the authors performed multiple rounds of selection to identify the important nucleocapsid features for enhanced genome packaging, nuclease protection, and circulation time *in vivo*, without compromising the architecture of the structure. This was the first reported case of a non-viral container that can encapsulate its own genome and evolve in a complex extracellular environment, with the synthetic systems serving to rival the best recombinant adeno-associated viruses. Using a similar approach, scientists used computational modeling to design and evolve proteins with different functions, including dimerization[132] and decreased side effects in a pre-clinical tumor model[133]. In particular, the authors \ designed a variant of interleukin-2 (IL-2) that would

bind its receptor on the target cell (T cells) without binding off-target receptors. The authors found that by redesigning one of the four helices on native IL-2 protein, they could increase on-target binding to the IL-2 receptor βγc heterodimer, while decreasing off-target binding to IL-2Rα (CD25), thereby driving toxicity. By redesigning these motifs, the authors improved IL-2 efficacy in mouse models of melanoma and colon cancer. Using a different approach, Guerette *et al.* coupled transcriptomics and proteomics data to design and predict the behavior of biomimetic materials[134]. The authors were able to rapidly process structural and functional novel high-performance eco-friendly materials pertaining to embryo protection, predation and adhesion. For example, they engineered silk-like materials from squid sucker ring teeth proteins that exceed the mechanical properties of many natural and synthetic polymers. Of particular note, the authors found a structural protein, suckerin-39, that surrounds squid sucker ring teeth and has high homology to silk, which would not have been discovered without the use of a combinatorial approach.

More recently, a series of papers have generated large biomaterial datasets without using NGS. In one example, quantitative structure-property relationship was retrospectively performed on a dataset describing nanoparticle formation; using this analysis, the authors found specific molecular variables associated with the drugs encapsulated in the nanoparticles were predictive of nanoparticle formation. Interestingly, the variables were related directly with the electronic configuration of the atoms making up the drug. Using only the molecular structure information of drug compounds, the authors rationally designed nanoparticles that delivered chemotherapeutics to tumors in mice[135], exploiting caveolin-dependent nanoparticle endocytosis. Specifically, the authors explored a number of different nanomaterial groups (e.g. detergents, azo dyes, and polyelectrolytes) and used their quantitative structure-nanoparticle assembly prediction model to predict, and then validate, whether 400 different hydrophobic drugs would formulate into nanoparticles. Taken together, these examples constitute an innovative approach to coupling computational techniques, experiments, and unbiased screens, in order to improve nanomaterial design. In a third example, Yamankurt et al. developed a high throughput method based on mass spectrometry to monitor how immune cells responded to spherical nucleic acid nanomedicines[136]. The authors designed a library of 960 nanomedicines, varying the nanoparticle core (e.g. cholesterol, phospholipid), oligonucleotide shell (e.g. phosphodiester or phosphorothioate backbone, and sequence), and peptide antigen (e.g. OVA or E7). Their high throughput cell toxicity assay led to several structure-function relationships. First, spherical nucleic acid nanomedicines elicit more immune activation than linear oligonucleotides, and linear oligonucleotide immune activation is dependent on what the oligonucleotide is conjugated to (e.g. cholesterol, DOPE) as well as its backbone. Notably, the authors used the data to ‘train’ a machine learning algorithm, in order to identify non-linear property interactions (e.g. if there are 5 different properties, what is the interdependent effect of each property on the other). This is important because it can be difficult to decouple the effect of one property on another in a high-throughput screen where lots of variables are being changed, thus making it challenging to predict the biological response to a nanomedicine. Most recently, Rath et al. released a pre-print describing VSEPRnet, a method by which the physical and chemical traits of biomolecules are encoded in a way that

enables neural network algorithms to make predictions[137]. The authors used this approach to predict binding between small peptides and allele-specific MHC-Class-1 molecules.

One need in the emerging field of large datasets and nanomedicine is the development of selection pressures that can be used to isolate nanoparticles that have performed a desired function *in vivo*. In biological studies, selection pressures are often based on cell death / proliferation, or alternatively, on fluorescence of a reporter gene[138–140]. High throughput nanotechnology screens will require assays with their own robust selection pressures, including biodistribution, functional cytoplasmic delivery, nuclear delivery, immunogenicity, and others. These will all generate different readouts. For example, nanoparticle delivery can be classified as (i) non-functional biodistribution, and (ii) functional, cytoplasmic delivery. In (i) a nanomaterial adhered to a cell is not distinguished from one that gets endocytosed, degraded in a lysosome, or delivered to the cytoplasm. However, in (ii) a nanoparticle must reach the cytoplasm of a cell, which ensures that only cells functionally delivered to are analyzed. These nanomaterial selection pressures can then be sub-divided into (i) up, or (ii) down-screens. Cells functionally delivered to in an up-screen change from no signal to a strong ‘on’ signal, whereas cells functionally delivered to in a down-screen change from high signal to ‘low’ signal.

Finally, well-designed studies could help answer key questions pertaining to the biology of delivery. First, which molecules play a predominant role in delivery? Proteins and lipids affect delivery, but carbohydrates require further exploration. Second, is a nanoparticle’s delivery more likely to be due to a small number of master regulatory genes, or many genes acting in concert? Third, do lncRNAs and epigenetic modifications alter the cellular response to nanoparticles? Given that these molecules regulate many biological phenotypes[5], we find it likely. Fourth, are there *in vitro* systems that efficiently recapitulate and predict *in vivo* delivery? Organ-on-chip systems may be poised to answer important biological questions. Finally, is there a ‘gold standard’ animal that can be used to predict delivery in large animals? The translation from delivery in small animal models (e.g. mice, rats) to efficient delivery in large animals (e.g. pigs, non-human primates, humans) is still largely unknown. The network analyses needed to answer these questions will be aided by multiomics. For example, sequencing technologies that concurrently measure mRNA expression and protein expression have been developed[141]. Multiomics analyses may also aid nanomedicines by improving the drugs nanomedicines are meant to deliver. For example, the efficacy of RNA therapies is strongly affected by chemical modifications to the RNA[142]. Transcriptomics can identify splicing patterns, as well as the frequency with which RNAs are affected by modifications. These modifications are known to affect maturation, folding, and metabolism[143–145] of mRNAs; understanding the relationship between modifications and RNA transport could lead to nucleic acid therapeutics with improved safety profiles.

The interface between materials, medicine, and high-throughput sequencing marks a significant opportunity for researchers. To take full advantage of novel technologies, nanotechnologists will need to understand molecular biology, data analysis, and data visualization. Currently, scientists who design nanoparticles do not typically work alongside scientists who study omics-sized data sets. One way to accelerate the marriage of omics and nanotechnology is to teach concepts like PCR, primer design, sequencing preparation, PCA,

and biostatistics in standard engineering and chemistry curricula. Until that time, if a chemist, materials scientists, or nanomedicine scientist would like to initiate an omics based experiment, it will be important to consider the following steps. First, identify the types of data that are necessary. Is it important to understand the transcriptomic response, epigenetic response, proteomic response, or some combination thereof? Is it sufficient to collect these data from many cells, or is it important to measure single cells individually? Second, seek out statisticians and bioinformaticians, in order to design your experiment correctly. How many groups or experimental conditions should be analyzed? What type of data analysis and visualization will be required? What types of experimental and technical controls are needed in order to believe the results? Answering these five questions will not guarantee the experiment is a success, but it will improve the odds that the data can be interpreted. Scientists who embrace NGS and analytics will be positioning themselves at the forefront of innovative new approaches that could accelerate the development of new materials and broadly benefit precision medicine and human health.

Acknowledgements.

The authors thanks John Platig at Harvard Medical School, Greg Gibson at Georgia Tech, Nirav Shaw at Georgia Tech, and Jordan E. Cattie at Emory University.

Funding. K.P., D.A.L., C.D.S., and J.E.D. were funded by Georgia Tech startup funds (awarded to J.E.D.) K.P. was funded by the NIH / NIGMS-sponsored Cell and Tissue Engineering (CTEng) Biotechnology Training Program (T32GM008433). C.D.S. was funded by the NIH Immunoengineering Training Program (T32). J.E.D. was funded by the Cystic Fibrosis Research Foundation (DAHLMA15XX0, awarded to J.E.D.), the Parkinson's Disease Foundation (PDF-JFA-1860, awarded to J.E.D.), and the Bayer Hemophilia Awards Program (AGE DTD, awarded to J.E.D.).

References

1. Hanahan D; Weinberg RA, Hallmarks of cancer: the next generation. Cell 2011, 144 (5), 646–74. [PubMed: 21376230]
2. Danial NN; Korsmeyer SJ, Cell death: critical control points. Cell 2004, 116 (2), 205–219. [PubMed: 14744432]
3. López-Otín C; Galluzzi L; Freije JM; Madeo F; Kroemer G, Metabolic control of longevity. Cell 2016, 166 (4), 802–821. [PubMed: 27518560]
4. Sorkin A; Von Zastrow M, Endocytosis and signalling: intertwining molecular networks. Nature reviews Molecular cell biology 2009, 10 (9), 609–622. [PubMed: 19696798]
5. Batista PJ; Chang HY, Long noncoding RNAs: cellular address codes in development and disease. Cell 2013, 152 (6), 1298–1307. [PubMed: 23498938]
6. Weinberg RA, Coming full circle-from endless complexity to simplicity and back again. Cell 2014, 157 (1), 267–71. [PubMed: 24679541]
7. Cheng CJ; Tietjen GT; Saucier-Sawyer JK; Saltzman WM, A holistic approach to targeting disease with polymeric nanoparticles. Nat Rev Drug Discov 2015, 14 (4), 239–47. [PubMed: 25598505]
8. Adams D; Gonzalez-Duarte A; O'Riordan WD; Yang CC; Ueda M; Kristen AV; Tournev I; Schmidt HH; Coelho T; Berk JL; Lin KP; Vita G; Attarian S; Plante-Bordeneuve V; Mezei MM; Campistol JM; Buades J; Brannagan TH 3rd; Kim BJ; Oh J; Parman Y; Sekijima Y; Hawkins PN; Solomon SD; Polydefkis M; Dyck PJ; Gandhi PJ; Goyal S; Chen J; Strahs AL; Nochur SV; Sweetser MT; Garg PP; Vaishnaw AK; Gollob JA; Suhr OB, Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. N. Engl. J. Med 2018, 379 (1), 11–21. [PubMed: 29972753]
9. Gilleron J; Querbes W; Zeigerer A; Borodovsky A; Marsico G; Schubert U; Manygoats K; Seifert S; Andree C; Stoter M; Epstein-Barash H; Zhang L; Koteliansky V; Fitzgerald K; Fava E; Bickle M; Kalaidzidis Y; Akinc A; Maier M; Zerial M, Image-based analysis of lipid nanoparticle-mediated

siRNA delivery, intracellular trafficking and endosomal escape. Nat. Biotechnol 2013, 31 (7), 638–46. [PubMed: 23792630]

10. Bahl K; Senn JJ; Yuzhakov O; Bulychev A; Brito LA; Hassett KJ; Laska ME; Smith M; Almarsson Ö; Thompson J, Preclinical and Clinical Demonstration of Immunogenicity by mRNA Vaccines against H10N8 and H7N9 Influenza Viruses. Mol. Ther 2017.
11. Ashton S; Song YH; Nolan J; Cadogan E; Murray J; Odedra R; Foster J; Hall PA; Low S; Taylor P; Ellston R; Polanska UM; Wilson J; Howes C; Smith A; Goodwin RJ; Swales JG; Strittmatter N; Takats Z; Nilsson A; Andren P; Trueman D; Walker M; Reimer CL; Troiano G; Parsons D; De Witt D; Ashford M; Hrkach J; Zale S; Jewsbury PJ; Barry ST, Aurora kinase inhibitor nanoparticles target tumors with favorable therapeutic index in vivo. Sci. Transl. Med 2016, 8 (325), 325ra17.
12. Caster JM; Patel AN; Zhang T; Wang A, Investigational nanomedicines in 2016: a review of nanotherapeutics currently undergoing clinical trials. Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology 2017, 9 (1).
13. Pasi KJ; Rangarajan S; Georgiev P; Mant T; Creagh MD; Lissitchkov T; Bevan D; Austin S; Hay CR; Hegemann I; Kazmi R; Chowdary P; Gercheva-Kyuchukova L; Mamonov V; Timofeeva M; Soh CH; Garg P; Vaishnaw A; Akinc A; Sorensen B; Ragni MV, Targeting of Antithrombin in Hemophilia A or B with RNAi Therapy. N. Engl. J. Med 2017, 377 (9), 819–828. [PubMed: 28691885]
14. Wilhelm S; Tavares AJ; Dai Q; Ohta S; Audet J; Dvorak HF; Chan WCW, Analysis of nanoparticle delivery to tumours. 2016, 1, 16014.
15. Leroux JC, Drug Delivery: Too Much Complexity, Not Enough Reproducibility? Angew. Chem. Int. Ed. Engl 2017.
16. Staff E, Time to Deliver. Nat. Biotechnol 2014, 32, 961. [PubMed: 25299892]
17. Harper AR; Topol EJ, Pharmacogenomics in clinical practice and drug development. Nat. Biotechnol 2012, 30 (11), 1117–24. [PubMed: 23138311]
18. Marx V, Biology: The big challenges of big data. Nature 2013, 498 (7453), 255–260. [PubMed: 23765498]
19. Tibbitt MW; Dahlman JE; Langer R, Emerging frontiers in drug delivery. J. Am. Chem. Soc 2016, 138 (3), 704–717. [PubMed: 26741786]
20. Cedervall T; Lynch I; Lindman S; Berggard T; Thulin E; Nilsson H; Dawson KA; Linse S, Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. Proc. Natl. Acad. Sci. U. S. A 2007, 104 (7), 2050–5. [PubMed: 17267609]
21. Wang F; Yu L; Monopoli MP; Sandin P; Mahon E; Salvati A; Dawson KA, The biomolecular corona is retained during nanoparticle uptake and protects the cells from the damage induced by cationic nanoparticles until degraded in the lysosomes. Nanomedicine 2013, 9 (8), 1159–68. [PubMed: 23660460]
22. Vilanova O; Mittag JJ; Kelly PM; Milani S; Dawson KA; Rädler JO; Franzese G, Understanding the kinetics of protein–nanoparticle corona formation. ACS nano 2016, 10 (12), 10842–10850. [PubMed: 28024351]
23. Raesch SS; Tenzer S; Storck W; Rurainski A; Selzer D; Ruge CA; Perez-Gil J; Schaefer UF; Lehr C-M, Proteomic and lipidomic analysis of nanoparticle corona upon contact with lung surfactant reveals differences in protein, but not lipid composition. ACS nano 2015, 9 (12), 11872–11885. [PubMed: 26575243]
24. Salvati A; Pitek AS; Monopoli MP; Prapainop K; Bombelli FB; Hristov DR; Kelly PM; Åberg C; Mahon E; Dawson KA, Transferrin-functionalized nanoparticles lose their targeting capabilities when a biomolecule corona adsorbs on the surface. Nat. Nanotechnol 2013, 8 (2), 137–143. [PubMed: 23334168]
25. Akinc A; Querbes W; De S; Qin J; Frank-Kamenetsky M; Jayaprakash KN; Jayaraman M; Rajeev KG; Cantley WL; Dorkin JR; Butler JS; Qin L; Racie T; Sprague A; Fava E; Zeigerer A; Hope MJ; Zerial M; Sah DW; Fitzgerald K; Tracy MA; Manoharan M; Koteliansky V; Fougerolles A; Maier MA, Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. Mol. Ther 2010, 18 (7), 1357–64. [PubMed: 20461061]

26. Zuckerman JE; Choi CH; Han H; Davis ME, Polycation-siRNA nanoparticles can disassemble at the kidney glomerular basement membrane. Proc. Natl. Acad. Sci. U. S. A 2012, 109 (8), 3137–42. [PubMed: 22315430]
27. Augustin HG; Koh GY, Organotypic vasculature: From descriptive heterogeneity to functional pathophysiology. Science 2017, 357 (6353), eaal2379. [PubMed: 28775214]
28. Tsoi KM; MacParland SA; Ma XZ; Spetzler VN; Echeverri J; Ouyang B; Fadel SM; Sykes EA; Goldaracena N; Kaths JM; Conneely JB; Alman BA; Selzner M; Ostrowski MA; Adeyi OA; Zilman A; McGilvray ID; Chan WC, Mechanism of hard-nanomaterial clearance by the liver. Nature materials 2016, 15 (11), 1212–1221. [PubMed: 27525571]
29. Sahay G; Querbes W; Alabi C; Eltoukhy A; Sarkar S; Zurenko C; Karagiannis E; Love K; Chen D; Zoncu R; Buganim Y; Schroeder A; Langer R; Anderson DG, Efficiency of siRNA delivery by lipid nanoparticles is limited by endocytic recycling. Nat. Biotechnol 2013, 31 (7), 653–8. [PubMed: 23792629]
30. Wittrup A; Ai A; Liu X; Hamar P; Trifonova R; Charisse K; Manoharan M; Kirchhausen T; Lieberman J, Visualizing lipid-formulated siRNA release from endosomes and target gene knockdown. Nat Biotechnol 2015, 33 (8), 870–6. [PubMed: 26192320]
31. Wang S; Sun H; Tanowitz M; Liang XH; Crooke ST, Annexin A2 facilitates endocytic trafficking of antisense oligonucleotides. Nucleic Acids Res. 2016, 44 (15), 7314–30. [PubMed: 27378781]
32. Linnane E; Davey P; Zhang P; Puri S; Edbrooke M; Chiarparin E; Revenko AS; Macleod AR; Norman JC; Ross SJ, Differential uptake, kinetics and mechanisms of intracellular trafficking of next-generation antisense oligonucleotides across human cancer cell lines. Nucleic Acids Res. 2019, 47 (9), 4375–4392. [PubMed: 30927008]
33. Patel S; Ashwanikumar N; Robinson E; DuRoss A; Sun C; Murphy-Benenato KE; Mihai C; Almarsson O; Sahay G, Boosting Intracellular Delivery of Lipid Nanoparticle-Encapsulated mRNA. Nano Lett. 2017, 17 (9), 5711–5718. [PubMed: 28836442]
34. Paunovska K; Gil CJ; Lokugamage MP; Sago CD; Sato M; Lando GN; Gamboa Castro M; Bryksin AV; Dahlman JE, Analyzing 2000 in Vivo Drug Delivery Data Points Reveals Cholesterol Structure Impacts Nanoparticle Delivery. ACS nano 2018, 12 (8), 8341–8349. [PubMed: 30016076]
35. Bertrand N; Grenier P; Mahmoudi M; Lima EM; Appel EA; Dormont F; Lim JM; Karnik R; Langer R; Farokhzad OC, Mechanistic understanding of in vivo protein corona formation on polymeric nanoparticles and impact on pharmacokinetics. Nature communications 2017, 8 (1), 777.
36. Sago CD; Lokugamage MP; Lando GN; Djeddar N; Shah NN; Syed C; Bryksin AV; Dahlman JE, Modifying a Commonly Expressed Endocytic Receptor Retargets Nanoparticles in Vivo. Nano Lett. 2018.
37. Lloyd-Evans E; Morgan AJ; He X; Smith DA; Elliot-Smith E; Sillence DJ; Churchill GC; Schuchman EH; Galione A; Platt FM, Niemann-Pick disease type C1 is a sphingosine storage disease that causes deregulation of lysosomal calcium. Nat. Med 2008, 14 (11), 1247–1255. [PubMed: 18953351]
38. Lifland AW; Jung J; Alonas E; Zurla C; Crowe JE; Santangelo PJ, Human respiratory syncytial virus nucleoprotein and inclusion bodies antagonize the innate immune response mediated by MDA5 and MAVS. J. Virol 2012, 86 (15), 8245–8258. [PubMed: 22623778]
39. Rohner NA; Thomas SN, Melanoma growth effects on molecular clearance from tumors and biodistribution into systemic tissues versus draining lymph nodes. J Control Release 2016, 223, 99–108. [PubMed: 26721446]
40. Wolfram J; Nizzero S; Liu H; Li F; Zhang G; Li Z; Shen H; Blanco E; Ferrari M, A chloroquine-induced macrophage-preconditioning strategy for improved nanodelivery. Sci. Rep 2017, 7 (1), 13738. [PubMed: 29062065]
41. Kishimoto TK; Ferrari JD; LaMothe RA; Kolte PN; Griset AP; O'Neil C; Chan V; Browning E; Chalishazar A; Kuhlman W, Improving the efficacy and safety of biologic drugs with tolerogenic nanoparticles. Nat. Nanotechnol 2016, 11 (10), 890–899. [PubMed: 27479756]
42. Albanese A; Tang PS; Chan WC, The effect of nanoparticle size, shape, and surface chemistry on biological systems. Annu. Rev. Biomed. Eng 2012, 14, 1–16. [PubMed: 22524388]

43. Vácha R; Martinez-Veracoechea FJ; Frenkel D, Receptor-mediated endocytosis of nanoparticles of various shapes. Nano Lett. 2011, 11 (12), 5391–5395. [PubMed: 22047641]
44. Sykes EA; Chen J; Zheng G; Chan WC, Investigating the impact of nanoparticle size on active and passive tumor targeting efficiency. 2014.
45. Kahvejian A; Quackenbush J; Thompson JF, What would you do if you could sequence everything? Nat. Biotechnol 2008, 26 (10), 1125. [PubMed: 18846086]
46. Navin N; Kendall J; Troge J; Andrews P; Rodgers L; McIndoo J; Cook K; Stepansky A; Levy D; Esposito D; Muthuswamy L; Krasnitz A; McCombie WR; Hicks J; Wigler M, Tumour evolution inferred by single-cell sequencing. Nature 2011, 472 (7341), 90–4. [PubMed: 21399628]
47. Tang F; Barbacioru C; Wang Y; Nordman E; Lee C; Xu N; Wang X; Bodeau J; Tuch BB; Siddiqui A; Lao K; Surani MA, mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009, 6 (5), 377–82. [PubMed: 19349980]
48. Patel AP; Tirosh I; Trombetta JJ; Shalek AK; Gillespie SM; Wakimoto H; Cahill DP; Nahed BV; Curry WT; Martuza RL, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 2014, 344 (6190), 1396–1401. [PubMed: 24925914]
49. Ting DT; Wittner BS; Ligorio M; Jordan NV; Shah AM; Miyamoto DT; Aceto N; Bersani F; Brannigan BW; Xega K, Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. Cell Rep. 2014, 8 (6), 1905–1918. [PubMed: 25242334]
50. Streets AM; Zhang X; Cao C; Pang Y; Wu X; Xiong L; Yang L; Fu Y; Zhao L; Tang F; Huang Y, Microfluidic single-cell whole-transcriptome sequencing. Proceedings of the National Academy of Sciences 2014, 111 (19), 7048–7053.
51. Wu AR; Neff NF; Kalisky T; Dalerba P; Treutlein B; Rothenberg ME; Mburu FM; Mantalas GL; Sim S; Clarke MF, Quantitative assessment of single-cell RNA-sequencing methods. Nat. Methods 2014, 11 (1), 41–46. [PubMed: 24141493]
52. Pérez-Torrado R; Rantsiou K; Perrone B; Navarro-Tapia E; Querol A; Cocolin L, Ecological interactions among Saccharomyces cerevisiae strains: insight into the dominance phenomenon. Sci. Rep 2017, 7, 43603. [PubMed: 28266552]
53. Meyer M; Reimand J; Lan X; Head R; Zhu X; Kushida M; Bayani J; Pressey JC; Lionel AC; Clarke ID, Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. Proceedings of the National Academy of Sciences 2015, 112 (3), 851–856.
54. Treutlein B; Lee QY; Camp JG; Mall M; Koh W; Shariati SAM; Sim S; Neff NF; Skotheim JM; Wernig M, Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature 2016, 534 (7607), 391–395. [PubMed: 27281220]
55. Zeisel A; Muñoz-Manchado AB; Codeluppi S; Lönnerberg P; La Manno G; Juréus A; Marques S; Munguba H; He L; Betsholtz C, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 2015, 347 (6226), 1138–1142. [PubMed: 25700174]
56. Usoskin D; Furlan A; Islam S; Abdo H; Lönnerberg P; Lou D; Hjerling-Leffler J; Haeggström J; Kharchenko O; Kharchenko PV, Unbiased classification of sensory neuron types by large-scale singlecell RNA sequencing. Nat. Neurosci 2015, 18 (1), 145–153. [PubMed: 25420068]
57. Villani A-C; Satija R; Reynolds G; Sarkizova S; Shekhar K; Fletcher J; Griesbeck M; Butler A; Zheng S; Lazo S, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 2017, 356 (6335), eaah4573. [PubMed: 28428369]
58. Villani AC; Shekhar K, Single-Cell RNA Sequencing of Human T Cells. Methods Mol. Biol 2017, 1514, 203–239. [PubMed: 27787803]
59. Papalexi E; Satija R, Single-cell RNA sequencing to explore immune cell heterogeneity. Nat. Rev. Immunol 2017.
60. Dixit A; Parnas O; Li B; Chen J; Fulco CP; Jerby-Arnon L; Marjanovic ND; Dionne D; Burks T; Raychowdhury R; Adamson B; Norman TM; Lander ES; Weissman JS; Friedman N; Regev A, Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 2016, 167 (7), 1853–1866.e17. [PubMed: 27984732]
61. Rozenblatt-Rosen O; Stubbington MJT; Regev A; Teichmann SA, The Human Cell Atlas: from vision to reality. Nature 2017, 550 (7677), 451–453. [PubMed: 29072289]

62. Dey SS; Kester L; Spanjaard B; Bienko M; van Oudenaarden A, Integrated genome and transcriptome sequencing of the same cell. Nat. Biotechnol 2015, 33, 285. [PubMed: 25599178]
63. Angermueller C; Clark SJ; Lee HJ; Macaulay IC; Teng MJ; Hu TX; Krueger F; Smallwood SA; Ponting CP; Voet T; Kelsey G; Stegle O; Reik W, Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat. Methods 2016, 13, 229. [PubMed: 26752769]
64. Cheow LF; Courtois ET; Tan Y; Viswanathan R; Xing Q; Tan RZ; Tan DSW; Robson P; Loh Y-H; Quake SR; Burkholder WF, Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. Nat. Methods 2016, 13, 833. [PubMed: 27525975]
65. Hou Y; Guo H; Cao C; Li X; Hu B; Zhu P; Wu X; Wen L; Tang F; Huang Y; Peng J, Singlecell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res. 2016, 26 (3), 304–19. [PubMed: 26902283]
66. Gjoneska E; Pfenning AR; Mathys H; Quon G; Kundaje A; Tsai LH; Kellis M, Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. Nature 2015, 518 (7539), 365–9. [PubMed: 25693568]
67. Zhang B; Wang J; Wang X; Zhu J; Liu Q; Shi Z; Chambers MC; Zimmerman LJ; Shaddox KF; Kim S, Proteogenomic characterization of human colon and rectal cancer. Nature 2014, 513 (7518), 382. [PubMed: 25043054]
68. Alfaro JA; Sinha A; Kislinger T; Boutros PC, Onco-proteogenomics: cancer proteomics joins forces with genomics. Nat. Methods 2014, 11 (11), 1107. [PubMed: 25357240]
69. Geyer PE; Kulak NA; Pichler G; Holdt LM; Teupser D; Mann M, Plasma proteome profiling to assess human health and disease. Cell systems 2016, 2 (3), 185–195. [PubMed: 27135364]
70. Mun DG; Bhin J; Kim S; Kim H; Jung JH; Jung Y; Jang YE; Park JM; Kim H; Jung Y; Lee H; Bae J; Back S; Kim SJ; Kim J; Park H; Li H; Hwang KB; Park YS; Yook JH; Kim BS; Kwon SY; Ryu SW; Park DY; Jeon TY; Kim DH; Lee JH; Han SU; Song KS; Park D; Park JW; Rodriguez H; Kim J; Lee H; Kim KP; Yang EG; Kim HK; Paek E; Lee S; Lee SW; Hwang D, Proteogenomic Characterization of Human Early-Onset Gastric Cancer. Cancer Cell 2019, 35 (1), 111–124.e10. [PubMed: 30645970]
71. Trapnell C; Cacchiarelli D; Grimsby J; Pokharel P; Li S; Morse M; Lennon NJ; Livak KJ; Mikkelsen TS; Rinn JL, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol 2014, 32 (4), 381. [PubMed: 24658644]
72. Yan J; Risacher SL; Shen L; Saykin AJ, Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. Briefings in bioinformatics 2017.
73. Carrow JK; Cross LM; Reese RW; Jaiswal MK; Gregory CA; Kaunas R; Singh I; Gaharwar AK, Widespread changes in transcriptome profile of human mesenchymal stem cells induced by two-dimensional nanosilicates. Proceedings of the National Academy of Sciences 2018, 115 (17), E3905E3913.
74. Feliu N; Kohonen P; Ji J; Zhang Y; Karlsson HL; Palmberg L; Nyström A; Fadeel B, Nextgeneration sequencing reveals low-dose effects of cationic dendrimers in primary human bronchial epithelial cells. ACS nano 2014, 9 (1), 146–163. [PubMed: 25530437]
75. Lucafò M; Gerdol M; Pallavicini A; Pacor S; Zorzet S; Da Ros T; Prato M; Sava G, Profiling the molecular mechanism of fullerene cytotoxicity on tumor cells by RNA-seq. Toxicology 2013, 314 (1), 183–192. [PubMed: 24125657]
76. Gliga AR; Edoff K; Caputo F; Källman T; Blom H; Karlsson HL; Ghibelli L; Traversa E; Ceccatelli S; Fadeel B, Cerium oxide nanoparticles inhibit differentiation of neural stem cells. Sci. Rep 2017, 7 (1), 9284. [PubMed: 28839176]
77. Simon DF; Domingos RF; Hauser C; Hutchins CM; Zerges W; Wilkinson KJ, RNA-Seq analysis of the effects of metal nanoparticle exposure on the transcriptome of Chlamydomonas reinhardtii. Appl. Environ. Microbiol 2013, AEM. 00998–13.
78. Beauvais-Flück R; Slaveykova VI; Cosio C, Transcriptomic and physiological responses of the green microalga Chlamydomonas reinhardtii during short-term exposure to subnanomolar methylmercury concentrations. Environ. Sci. Technol 2016, 50 (13), 7126–7134. [PubMed: 27254783]

79. Zheng M; Lu J; Zhao D, Toxicity and Transcriptome Sequencing (RNA-seq) Analyses of Adult Zebrafish in Response to Exposure Carboxymethyl Cellulose Stabilized Iron Sulfide Nanoparticles. Sci. Rep 2018, 8 (1), 8083. [PubMed: 29795396]
80. Veiseh O; Doloff JC; Ma M; Vegas AJ; Tam HH; Bader AR; Li J; Langan E; Wyckoff J; Loo WS; Jhunjhunwala S; Chiu A; Siebert S; Tang K; Hollister-Lock J; Aresta-Dasilva S; Bochenek M; Mendoza-Elias J; Wang Y; Qi M; Lavin DM; Chen M; Dholakia N; Thakrar R; Lacik I; Weir GC; Oberholzer J; Greiner DL; Langer R; Anderson DG, Size- and shape-dependent foreign body immune response to materials implanted in rodents and non-human primates. Nature materials 2015, 14 (6), 643–51. [PubMed: 25985456]
81. Hebels D; Carlier A; Coonen MLJ; Theunissen DH; de Boer J, cBiT: A transcriptomics database for innovative biomaterial engineering. Biomaterials 2017, 149, 88–97. [PubMed: 29020642]
82. Conesa A; Madrigal P; Tarazona S; Gomez-Cabrero D; Cervera A; McPherson A; Szcześniak MW; Gaffney DJ; Elo LL; Zhang X; Mortazavi A, A survey of best practices for RNA-seq data analysis. Genome Biol. 2016, 17 (1), 13. [PubMed: 26813401]
83. Ronan T; Qi Z; Naegle KM, Avoiding common pitfalls when clustering biological data. Sci. Signaling 2016, 9 (432), re6–re6.
84. Novembre J; Stephens M, Interpreting principal component analyses of spatial population genetic variation. Nat. Genet 2008, 40 (5), 646–9. [PubMed: 18425127]
85. Ringnér M, What is principal component analysis? Nat. Biotechnol 2008, 26 (3), 303–304. [PubMed: 18327243]
86. Stegle O; Parts L; Piipari M; Winn J; Durbin R, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc 2012, 7 (3), 500–7. [PubMed: 22343431]
87. Mecham BH; Nelson PS; Storey JD, Supervised normalization of microarrays. Bioinformatics 2010, 26 (10), 1308–15. [PubMed: 20363728]
88. Lever J; Krzywinski M; Altman N, Points of Significance: Principal component analysis. Nat. Methods 2017, 14 (7), 641–642.
89. Maaten L v. d.; Hinton, G., Visualizing data using t-SNE. Journal of machine learning research 2008, 9 (Nov), 2579–2605.
90. Habib N; Avraham-Davidi I; Basu A; Burks T; Shekhar K; Hofree M; Choudhury SR; Aguet F; Gelfand E; Ardlie K, Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 2017, 14 (10), 955. [PubMed: 28846088]
91. Macosko EZ; Basu A; Satija R; Nemesh J; Shekhar K; Goldman M; Tirosh I; Bialas AR; Kamitaki N; Martersteck EM, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 2015, 161 (5), 1202–1214. [PubMed: 26000488]
92. Shalek AK; Satija R; Shuga J; Trombetta JJ; Gennert D; Lu D; Chen P; Gertner RS; Gaublomme JT; Yosef N; Schwartz S; Fowler B; Weaver S; Wang J; Wang X; Ding R; Raychowdhury R; Friedman N; Hacohen N; Park H; May AP; Regev A, Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 2014, 510 (7505), 363–9. [PubMed: 24919153]
93. Wattenberg M. a. V., Fernanda and Johnson Ian, How to Use t-SNE Effectively. Distill 2016.
94. Ester M; Kriegel H-P; Sander J; Xu X In A density-based algorithm for discovering clusters in large spatial databases with noise, Kdd, 1996; pp 226–231.
95. Kohonen T, The self-organizing map. Neurocomputing 1998, 21 (1–3), 1–6.
96. Hartigan JA; Wong MA, Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 1979, 28 (1), 100–108.
97. Wiwie C; Baumbach J; Röttger R, Comparing the performance of biomedical clustering methods. Nat. Methods 2015, 12 (11), 1033–1038. [PubMed: 26389570]
98. Xu C; Su Z, Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics 2015, 31 (12), 1974–1980. [PubMed: 25805722]
99. Jiang D; Tang C; Zhang A, Cluster analysis for gene expression data: A survey. IEEE Transactions on knowledge and data engineering 2004, 16 (11), 1370–1386.
100. Quackenbush J, Computational genetics: computational analysis of microarray data. Nature reviews genetics 2001, 2 (6), 418.

101. Jain AK, Data clustering: 50 years beyond K-means. Pattern Recogn. Lett 2010, 31 (8), 651–666.
102. Paunovska K; Sago CD; Monaco CM; Hudson WH; Castro MG; Rudoltz TG; Kalathoor S; Vanover DA; Santangelo PJ; Ahmed R; Bryksin AV; Dahlman JE, A Direct Comparison of in Vitro and in Vivo Nucleic Acid Delivery Mediated by Hundreds of Nanoparticles Reveals a Weak Correlation. Nano Lett. 2018, 18 (3), 2148–2157. [PubMed: 29489381]
103. Ding C; He X In K-nearest-neighbor consistency in data clustering: incorporating local information into global optimization, Proceedings of the 2004 ACM symposium on Applied computing, ACM: 2004; pp 584–589.
104. Handl J; Knowles J; Kell DB, Computational cluster validation in post-genomic data analysis. Bioinformatics 2005, 21 (15), 3201–3212. [PubMed: 15914541]
105. Gehlenborg N; O'donoghue SI; Baliga NS; Goesmann A; Hibbs MA; Kitano H; Kohlbacher O; Neuweger H; Schneider R; Tenenbaum D, Visualization of omics data for systems biology. Nat. Methods 2010, 7 (3s), S56. [PubMed: 20195258]
106. Kanehisa M; Furumichi M; Tanabe M; Sato Y; Morishima K, KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017, 45 (D1), D353–d361. [PubMed: 27899662]
107. Kanehisa M; Goto S, KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000, 28 (1), 27–30. [PubMed: 10592173]
108. Kanehisa M; Sato Y; Kawashima M; Furumichi M; Tanabe M, KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016, 44 (D1), D457–62. [PubMed: 26476454]
109. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 2017, 45 (D1), D331–d338. [PubMed: 27899567]
110. Ashburner M; Ball CA; Blake JA; Botstein D; Butler H; Cherry JM; Davis AP; Dolinski K; Dwight SS; Eppig JT; Harris MA; Hill DP; Issel-Tarver L; Kasarskis A; Lewis S; Matese JC; Richardson JE; Ringwald M; Rubin GM; Sherlock G, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet 2000, 25 (1), 25–9. [PubMed: 10802651]
111. Shannon P; Markiel A; Ozier O; Baliga NS; Wang JT; Ramage D; Amin N; Schwikowski B; Ideker T, Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003, 13 (11), 2498–2504. [PubMed: 14597658]
112. Tong AHY; Evangelista M; Parsons AB; Xu H; Bader GD; Pagé N; Robinson M; Raghibizadeh S; Hogue CW; Bussey H, Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 2001, 294 (5550), 2364–2368. [PubMed: 11743205]
113. Subramanian A; Narayan R; Corsello SM; Peck DD; Natoli TE; Lu X; Gould J; Davis JF; Tubelli AA; Asiedu JK, A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 2017, 171 (6), 1437–1452. e17. [PubMed: 29195078]
114. Hughes TR; Marton MJ; Jones AR; Roberts CJ; Stoughton R; Armour CD; Bennett HA; Coffey E; Dai H; He YD, Functional discovery via a compendium of expression profiles. Cell 2000, 102 (1), 109–126. [PubMed: 10929718]
115. Dahlman JE; Kauffman KJ; Xing Y; Shaw TE; Mir FF; Dlott CC; Langer R; Anderson DG; Wang ET, Barcoded nanoparticles for high throughput *in vivo* discovery of targeted therapeutics. Proc. Natl. Acad. Sci. U. S. A 2017, 114 (8), 2060–2065. [PubMed: 28167778]
116. Wong B, Points of view: Color coding. Nat. Methods 2010, 7 (8), 573–573. [PubMed: 20704014]
117. Hansen KD; Brenner SE; Dudoit S, Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 2010, 38 (12), e131–e131. [PubMed: 20395217]
118. Loman NJ; Misra RV; Dallman TJ; Constantinidou C; Gharbia SE; Wain J; Pallen MJ, Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol 2012, 30 (5), 434. [PubMed: 22522955]
119. Ross MG; Russ C; Costello M; Hollinger A; Lennon NJ; Hegarty R; Nusbaum C; Jaffe DB, Characterizing and measuring bias in sequence data. Genome Biol. 2013, 14 (5), R51. [PubMed: 23718773]
120. Asmann YW; Klee EW; Thompson EA; Perez EA; Middha S; Oberg AL; Therneau TM; Smith DI; Poland GA; Wieben ED, 3'tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. BMC Genomics 2009, 10 (1), 531. [PubMed: 19917133]

121. Griffith M; Griffith OL; Mwenifumbo J; Goya R; Morrissy AS; Morin RD; Corbett R; Tang MJ; Hou Y-C; Pugh TJ, Alternative expression analysis by RNA sequencing. Nat. Methods 2010, 7 (10), 843. [PubMed: 20835245]

122. Sago CD; Lokugamage MP; Paunovska K; Vanover DA; Monaco CM; Shah NN; Gamboa Castro M; Anderson SE; Rudoltz TG; Lando GN; Mummilal Tiwari P; Kirschman JL; Willett N; Jang YC; Santangelo PJ; Bryksin AV; Dahlman JE, High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. Proceedings of the National Academy of Sciences 2018.

123. Sago CD; Lokugamage MP; Islam FZ; Krupczak BR; Sato M; Dahlman JE, Nanoparticles that deliver RNA to bone marrow identified by in vivo directed evolution. J. Am. Chem. Soc 2018.

124. Lokugamage MP; Sago CD; Dahlman JE, Testing thousands of nanoparticles in vivo using DNA barcodes. Current Opinion in Biomedical Engineering 2018.

125. Paunovska K; Da Silva Sanchez AJ; Sago CD; Gan Z; Lokugamage MP; Islam FZ; Kalathoor S; Krupczak BR; Dahlman JE, Nanoparticles Containing Oxidized Cholesterol Deliver mRNA to the Liver Microenvironment at Clinically Relevant Doses. Adv Mater 2019, e1807748. [PubMed: 30748040]

126. Yaari Z; da Silva D; Zinger A; Goldman E; Kajal A; Tshuva R; Barak E; Dahan N; Hershkovitz D; Goldfeder M; Roitman JS; Schroeder A, Theranostic barcoded nanoparticles for personalized cancer medicine. Nature communications 2016, 7, 13325.

127. Bhatia SN; Ingber DE, Microfluidic organs-on-chips. Nat. Biotechnol 2014, 32 (8), 760–72. [PubMed: 25093883]

128. Chan KY; Jang MJ; Yoo BB; Greenbaum A; Ravi N; Wu WL; Sanchez-Guardado L; Lois C; Mazmanian SK; Deverman BE; Gradinaru V, Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci 2017, 20 (8), 1172–1179. [PubMed: 28671695]

129. Deverman BE; Pravdo PL; Simpson BP; Kumar SR; Chan KY; Banerjee A; Wu WL; Yang B; Huber N; Pasca SP; Gradinaru V, Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol 2016, 34 (2), 204–9. [PubMed: 26829320]

130. Belser JA; Katz JM; Tumpey TM, The ferret as a model organism to study influenza A virus infection. Dis. Model. Mech 2011, 4 (5), 575–9. [PubMed: 21810904]

131. Butterfield GL; Lajoie MJ; Gustafson HH; Sellers DL; Nattermann U; Ellis D; Bale JB; Ke S; Lenz GH; Yehdego A, Evolution of a designed protein assembly encapsulating its own RNA genome. Nature 2017, 552 (7685), 415. [PubMed: 29236688]

132. Chen Z; Boyken SE; Jia M; Busch F; Flores-Solis D; Bick MJ; Lu P; VanAernum ZL; Sahasrabuddhe A; Langan RA; Bermeo S; Brunette TJ; Mulligan VK; Carter LP; DiMaio F; Sgourakis NG; Wysocki VH; Baker D, Programmable design of orthogonal protein heterodimers. Nature 2019, 565 (7737), 106–111. [PubMed: 30568301]

133. Silva DA; Yu S; Ulge UY; Spangler JB; Jude KM; Labao-Almeida C; Ali LR; Quijano-Rubio A; Ruterbusch M; Leung I; Biary T; Crowley SJ; Marcos E; Walkey CD; Weitzner BD; Pardo-Avila F; Castellanos J; Carter L; Stewart L; Riddell SR; Pepper M; Bernardes GJL; Dougan M; Garcia KC; Baker D, De novo design of potent and selective mimics of IL-2 and IL-15. Nature 2019, 565 (7738), 186–191. [PubMed: 30626941]

134. Guerette PA; Hoon S; Seow Y; Raida M; Masic A; Wong FT; Ho VH; Kong KW; Demirel MC; Pena-Francesch A; Amini S; Tay GZ; Ding D; Miserez A, Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. Nat. Biotechnol 2013, 31 (10), 908–15. [PubMed: 24013196]

135. Shamay Y; Shah J; Isik M; Mizrachi A; Leibold J; Tschaharganeh DF; Roxbury D; Budhathoki-Uprety J; Nawaly K; Sugarman JL; Baut E; Neiman MR; Dacek M; Ganesh KS; Johnson DC; Sridharan R; Chu KL; Rajasekhar VK; Lowe SW; Chodera JD; Heller DA, Quantitative self-assembly prediction yields targeted nanomedicines. Nature materials 2018, 17 (4), 361368.

136. Yamankurt G; Berns EJ; Xue A; Lee A; Bagheri N; Mrksich M; Mirkin CA, Exploration of the nanomedicine-design space with high-throughput screening and machine learning. Nature biomedical engineering 2019, 3 (4), 318–327.

137. Rath S; Francis-Landau J; Lu X; Nakano-Baker O; Rodriguez J; Ustundag BB; Sarikaya M, VSEPRnet: Physical structure encoding of sequence-based biomolecules for functionality prediction: Case study with peptides. bioRxiv 2019, 656033.

138. Shalem O; Sanjana NE; Zhang F, High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet 2015, 16 (5), 299–311. [PubMed: 25854182]

139. Chen S; Sanjana NE; Zheng K; Shalem O; Lee K; Shi X; Scott DA; Song J; Pan JQ; Weissleder R; Lee H; Zhang F; Sharp PA, Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell 2015, 160 (6), 1246–60. [PubMed: 25748654]

140. Shalem O; Sanjana NE; Hartenian E; Shi X; Scott DA; Mikkelsen TS; Heckl D; Ebert BL; Root DE; Doench JG; Zhang F, Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014, 343 (6166), 84–7. [PubMed: 24336571]

141. Darmanis S; Gallant CJ; Marinescu VD; Niklasson M; Segerman A; Flamourakis G; Fredriksson S; Assarsson E; Lundberg M; Nelander S, Simultaneous multiplexed measurement of RNA and proteins in single cells. Cell Rep. 2016, 14 (2), 380–389. [PubMed: 26748716]

142. Foster DJ; Brown CR; Shaikh S; Trapp C; Schlegel MK; Qian K; Sehgal A; Rajeev KG; Jadhav V; Manoharan M; Kuchimanchi S; Maier MA; Milstein S, Advanced siRNA Designs Further Improve In Vivo Performance of GalNAc-siRNA Conjugates. Mol. Ther 2018, 26 (3), 708–717. [PubMed: 29456020]

143. Vu LP; Pickering BF; Cheng Y; Zaccara S; Nguyen D; Minuesa G; Chou T; Chow A; Saletore Y; MacKay M; Schulman J; Famulare C; Patel M; Klimek VM; Garrett-Bakelman FE; Melnick A; Carroll M; Mason CE; Jaffrey SR; Kharas MG, The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. Nat. Med 2017, 23 (11), 1369–1376. [PubMed: 28920958]

144. Roundtree IA; Evans ME; Pan T; He C, Dynamic RNA Modifications in Gene Expression Regulation. Cell 2017, 169 (7), 1187–1200. [PubMed: 28622506]

145. Zhao BS; Roundtree IA; He C, Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. 2017, 18 (1), 31–42. [PubMed: 27808276]

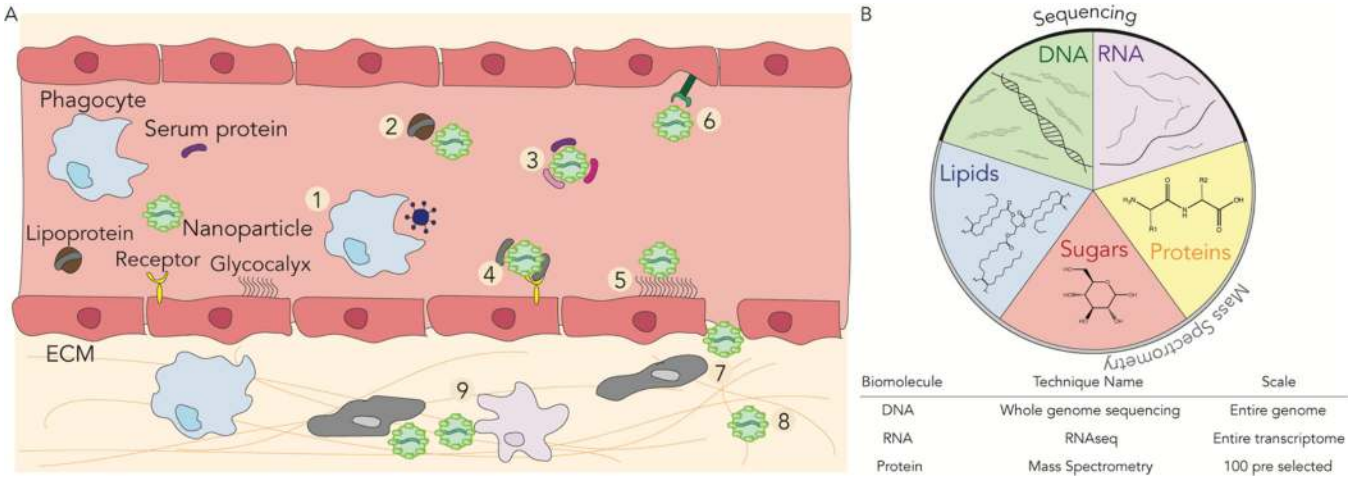

Figure 1.

Nanoparticle delivery can be viewed as a complex phenotype affected by many cells and biomolecules. **(A)** Nanoparticles are (1) cleared by circulating immune cells and tissue resident immune cells. Due to their high surface area: volume ratio, nanoparticles interface with (2) lipoproteins and (3) other biomolecules that make up the protein corona. The corona, in turn, can (4) alter how nanoparticles bind target cells. Interestingly, depending on its composition, the nanoparticle corona can promote or inhibit cell targeting. While reaching target cells, nanoparticles also interact with (5) a dense 'forest' of cell surface glycoproteins and glycolipids, collectively termed the glycocalyx. Alternatively, nanoparticles may interact (6) directly with cell surface receptors. Nanoparticles can also exit the bloodstream; this process is affected by (7) the permeability of vascular endothelial cells. Within the target tissue, nanoparticles interact with (8) proteoglycans in the extracellular matrix (ECM), or (9) cells within the tissue itself. **(B)** DNA- and RNA-driven gene expression dictates nanoparticle behavior by controlling the synthesis and processing of proteins, sugars, and lipids. As a result, high throughput quantification of the 5 biomolecules could improve our understanding of biological pathways that affect nanoparticle delivery. Two methods are typically used: next generation sequencing, which quantifies DNA and RNA, and mass spectroscopy, which quantifies lipids, carbohydrates, and proteins. The scale at which DNA and RNA can be analyzed is currently greater than the scale at which lipids, carbohydrates, and proteins can be analyzed.

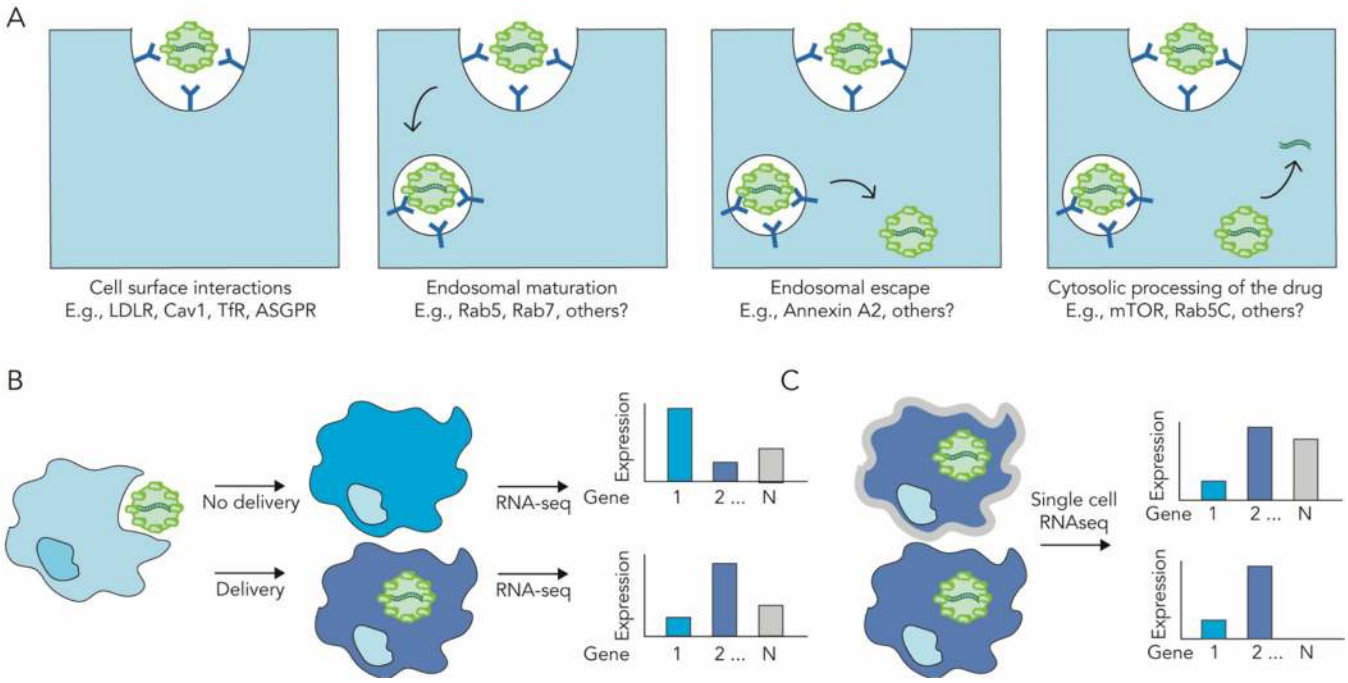

Figure 2.

Transcriptomics can be used to study how cell respond to nanomaterials. (A) Gene expression can alter how nanoparticles interact with the cell surface, how endosomes mature, how nanoparticles are released from the endosome, and how the drug is processed after it is delivered into the cytoplasm. (B) To measure gene expression changes caused by nanomaterials, cells that do (and do not) uptake nanoparticles can be separated. Using RNA-seq to compare these two populations of cells, individual genes and pathways to promote or prevent delivery can be identified. (C) Single cell RNA-seq (scRNAseq) may identify subpopulations of cells that respond to nanoparticles in a unique way. In this example, when analyzed with scRNA-seq, the expression of gene 1 and 2 does not change, relative to the analysis of many cells (depicted in (B)). By contrast, the expression of gene N varies significantly across individual cells in a way that cannot be quantified using bulk analysis.

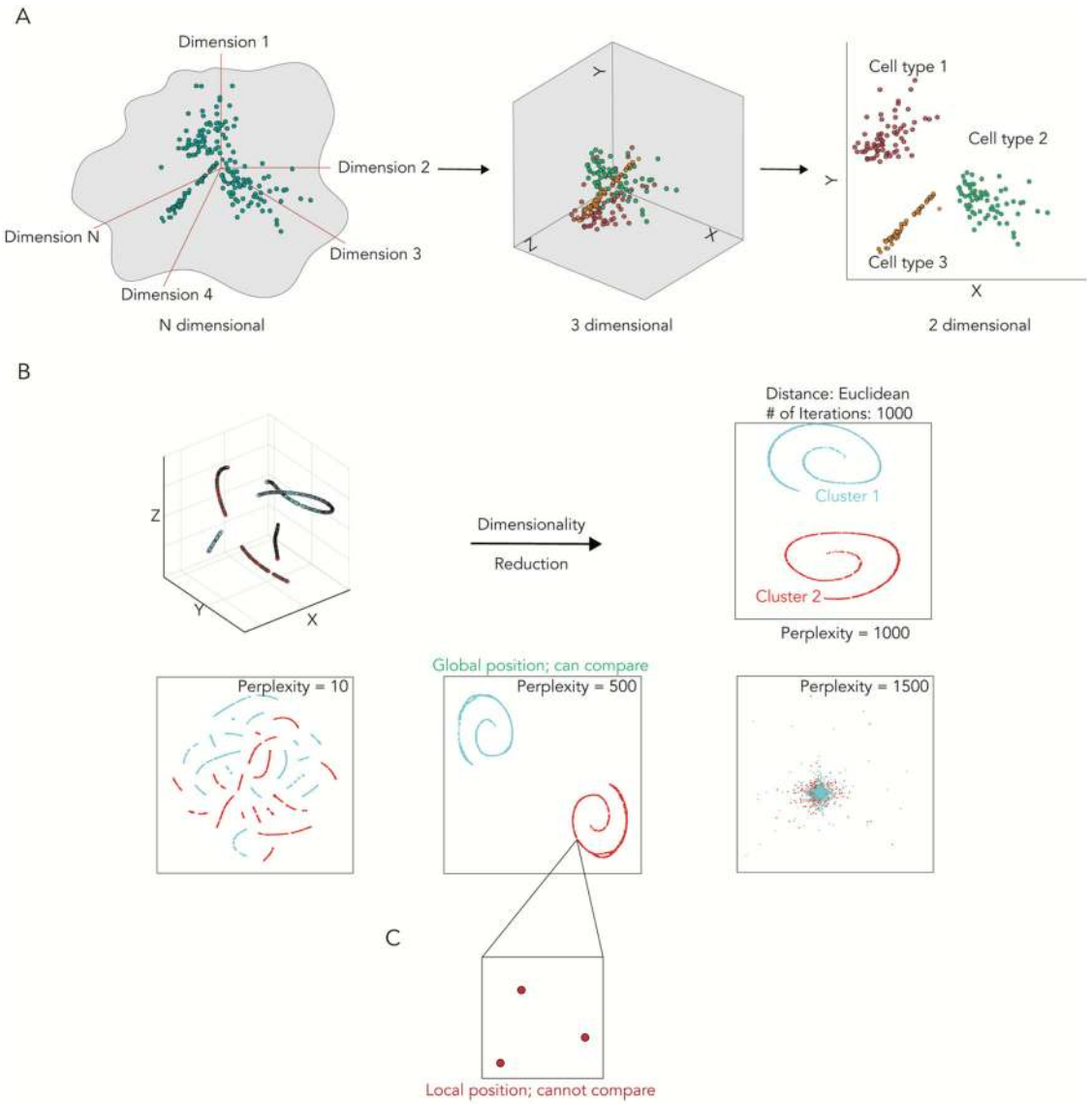

Figure 3.

After generating large datasets, (A) data can be reduced to a smaller number of dimensions. This is done so data can be clearly visualized after identifying the most important variables in the experiment. (B) When reducing data dimensionality, selecting incorrect input variables can lead to images that contain clustered data when no clusters actually exist. In this example, varying the perplexity variable alters clustering. (C) Interpreting relationships between individual points in a t-SNE plot is not appropriate since the position of individual dots varies with each run of the analysis. Interpreting broad relationships from the data is appropriate.

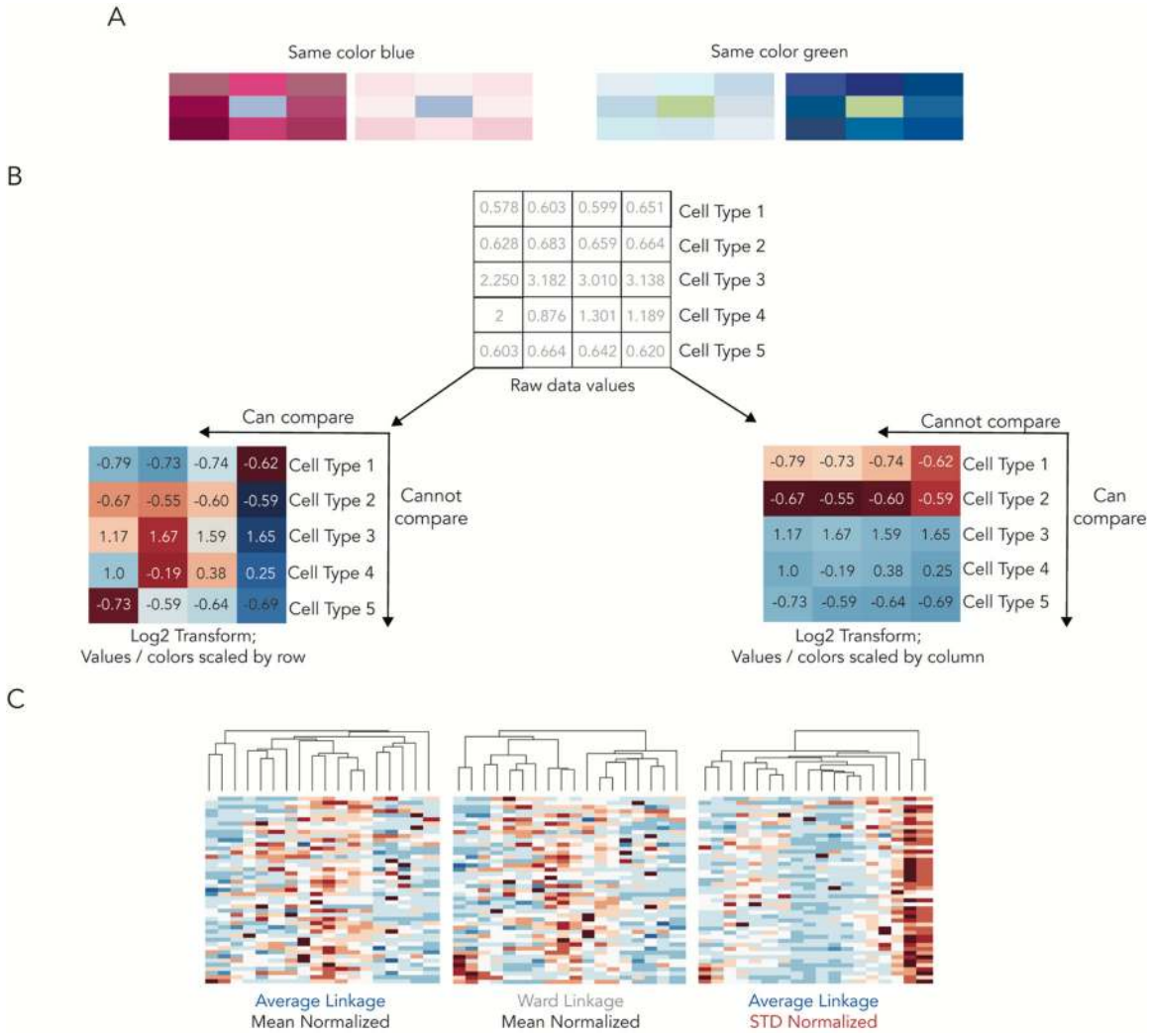

Figure 4.

Heatmap generation and interpretation depends on the algorithms, conditions, and colors used. (A) The same color can look different when surrounded by different colors. (B) Heatmaps can be scaled by row or column. If scaling by row, colors can be compared within a row. If scaling by column, colors can be compared within the column. (D) Dendrogram clusters vary as a function of the normalization method and clustering algorithms.

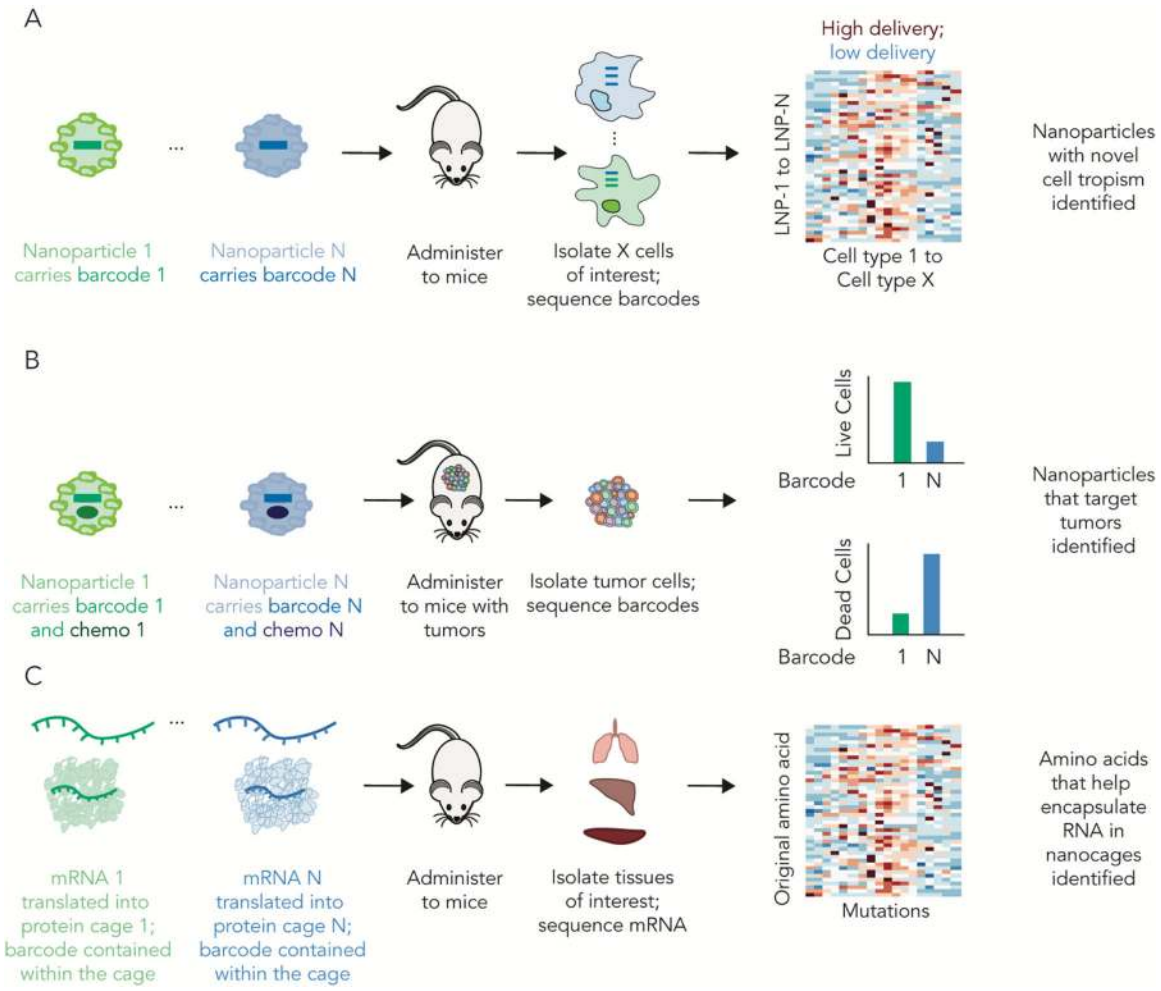

Figure 5.

High throughput *in vivo* assays have been used to study nanomedicines. (A) In one example, nanoparticles were formulated to carry DNA barcodes. Nanoparticle 1, with chemical structure 1, was made to carry DNA barcode 1; nanoparticle N, with chemical structure N, was made to carry DNA barcode N. All N nanoparticles were administered to mice, cells of interest were isolated, and next generation sequencing was using to quantify delivery of all N nanoparticles simultaneously. (B) In another example, liposome 1 was formulated to carry DNA barcode 1 and a chemotherapy; liposome N was formulated to carry DNA barcode N and a chemotherapy. Tumor delivery was quantified by measuring live / dead cells isolated from the tumor. (C) In a third example, nanocages consisting of a different protein shell were encoded with mRNAs. The protein nanocages were administered to mice, and the effective nanocages were isolated from tissues. Sequencing was used to determine the mRNAs, and thus, by extension, the protein nanocages that survived *in vivo*.