Abstract
Nanoparticles improve drug efficacy by delivering drugs to sites of disease. To effectively deliver a drug in vivo, a nanoparticle must overcome physical and physiological hurdles that are not present in cell culture, yet in vitro screens are used to predict nanoparticle delivery in vivo. An ideal nanoparticle discovery pipeline would enable scientists to study thousands of nanoparticles in vivo. Here, we discuss technologies that enable high throughput in vitro screens, focusing on DNA barcoded nanoparticles.

Addresses
Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory School of Medicine, Atlanta, GA, USA

Corresponding author: Dahlman, James E. (james.dahlman@bme.gatech.edu)

RNA therapies are limited by inefficient delivery to target tissues
Advances in genomics have armed scientists with lists of genes that cause diseases. This has brought about a significant change in the way drugs are designed and discovered. Researchers were previously limited to targeting broad cellular phenotypes; for example, cisplatin intercalates into double stranded DNA, causing toxicity in any cell undergoing cell division. Although many of these drugs successfully treated disease, they also caused severe side effects driven by drug activity in ‘off-target’ cells. Researchers now often use small molecules that target specific mutations. However, only 15% of the protein coding genome — and a much smaller percentage of the non-coding genome — is ‘druggable’ using small molecules [1]. This has led scientists to develop technologies to target all genes.

RNA therapies have emerged as a promising solution to the problem of ‘undruggable’ targets [2]. RNAs can specifically turn any gene in the genome on or off, regardless of its eventual protein structure. Gene silencing can be mediated by RNA interference (RNAi), a well characterized mechanism where base pairing between small interfering RNA (siRNA) and target mRNA catalyzes RISC-mediated mRNA degradation. Gene silencing can also be mediated by DNA nucleases, including those derived from zinc fingers (ZFNs) [3], transcription factor-like effectors (TALENs) [4], or clustered regularly interspaced short palindromic repeats and their associated proteins (CRISPR-Cas) systems [5,6]. Similarly, mRNA-based therapies can transiently upregulate genes [7]. To date siRNA therapies have shown the most promise in patients. In one example, a degenerative disease that was uniformly fatal has been halted (and in some cases, reversed) in Phase III clinical trials [8,9]. More generally, over 1200 patients have been treated with siRNA, some for as long as 36 months. With a few exceptions, primarily limited to antisense nucleotides (which are distinct from siRNA) [10], siRNA therapies have been well tolerated and efficacious. Advances in siRNA biochemistry [11] and improved delivery vehicles [12] raise the exciting prospect that once yearly injections to treat genetic disease are within reach.

Although siRNA is currently the most clinically advanced RNA therapy, DNA nucleases and mRNA therapies are up and coming. They have shown efficacy in mice [13,14] and non-human primates [15], and are poised to treat disease in humans [16–18]. The results described above have one limitation: most clinical trials have involved RNAs locally injected into muscle (e.g., vaccines), eye, or lymph nodes, administered to cells ex vivo, or systemically delivered to hepatocytes. Systemically delivering therapeutic RNA outside the liver remains a substantial unsolved problem [19] that limits the development of gene therapies targeting other organs. RNA therapies will require delivery because naked RNAs are quickly degraded by nucleases, and their large molecular weight and highly negative phosphodiester backbone prevents them from crossing the anionic cell membrane [20].

Thousands of nanoparticles are screened in vitro; this may not predict in vivo delivery
RNA delivery vehicles are designed to protect the nucleic acid and transport it to the target cell. Although
many drug delivery vehicles have been used to deliver RNA, we will focus on nanoparticles, which have generated promising clinical data [8]. Here we define a nanoparticle as a structure with all 3 dimensions less than 1000 nm. Scientists have made steady advances in nanoparticle design. Nanoparticles can be made with variable size [21], ionizability [22], hydrophilicity [23], shape [24], and with varying degrees of active targeting.

Figure 1

(A) Lipid nanoparticle libraries can be generated with a variety of chemical compounds. LNPs are formulated by combining different biomaterials with cholesterols, lipid-PEG compounds, and helper lipids. (B) High-throughput nanoparticle formulation allows for the rapid production of large, diverse libraries. (C) Nanoparticle libraries can be generated by varying the molar ratio of biomaterials, cholesterols, PEG, and helper lipids. (D) Between $1 \times 10^8$ and $2 \times 10^{11}$ chemically distinct nanoparticles can be made by combining these compounds. Notably, these numbers do not include variations in the molar ratio of the compounds (e.g., Figure 1C).
ligands [25]. Large, chemically diverse libraries have been synthesized using simple synthetic routes including, but not limited to, Michael addition [26], epoxides [27], peptide [28], and thiol chemistry [29]. Importantly, advances in nanoparticle formulation — defined here as the process of ‘loading’ the nucleic acid into the nanoparticle — have also been reported. High throughput microfluidics has been shown to reliably make small, consistent batches of nanoparticles that are stable for weeks [30,31]. It is still difficult to cover the entire nanoparticle chemical space; formulating nanoparticles using available chemistries, it is feasible to formulate between 100 million and 200 billion chemically distinct nanoparticles (Figure 1).

After nanoparticles are formulated, they are screened in vitro. More specifically, scientists (a) synthesize thousands of different nanoparticles before (b) evaluating whether they deliver RNA in easily expandable cell lines (e.g., HeLa). Scientists have also used primary cells in place of immortalized cells [32]. Based on in vitro results, a very small number of nanoparticles are (c) tested in vivo. This process is only an efficient use of time and resources if in vitro nanoparticle delivery predicts in vivo delivery. To test this, we compared how the same 300 nanoparticles delivered nucleic acids in vitro and in vivo in multiple cell types; we found no correlation [33] (Figure 2A). The discrepancy between in vitro and in vivo delivery is not necessarily surprising. If a nanoparticle is systemically administered, it must travel through the blood stream and enter the target tissue. Several physical factors dictate where the nanoparticles go. For example, it is difficult for nanoparticles to access the brain, which is physically cordoned off by the blood brain barrier. Nanoparticles can also be physically disassembled by the glomerular membrane in the kidney [34]. As a counterexample, nanoparticles often target hepatocytes since nanoparticles can exit the blood via nanoscale pores in sinusoidal blood vessels and basement membrane [35]. Physiological factors also influence nanoparticle delivery. For example, serum proteins can bind to nanoparticles in the blood [36,37], changing nanoparticle interactions with the immune system and target cells [38]. Interestingly, systemic nanoparticle delivery can also change with local disease states; in one example, scientists found that delivery to systemic organs was affected by the presence of a primary tumor [39]. Even if the nanoparticle reaches its target cell, it must enter the cytoplasm, often by escaping an endosome. Endosomal escape is inefficient; a LNP that delivers siRNA to hepatocytes very efficiently (50% target gene silencing after a 0.01 mg/kg injection) still had >95% of its siRNA sequestered within endosomes [40]. The nanoparticle must also evade liver, kidney, and splenic clearance, as well as avoid initiating an immune response. The majority of these obstacles are not recapitulated in vitro. Some processes are required for in vitro delivery and in vivo delivery (e.g., endocytosis and endosomal escape). However, these processes are carefully regulated [41] by gene expression that is likely to change with microenvironmental cues.

Figure 2

DNA barcodes have analyzed nanoparticle delivery in vivo. (A) Chemically diverse LNPs are tagged with unique DNA barcodes. LNPs were pooled together and administered both in vitro and in vivo. For in vitro studies, delivery at condition 1 often correlates with delivery at condition 2. However, in vitro delivery does not correlate to in vivo delivery. These results were consistent across many cell types, and across different experiments, for >300 different LNPs [33]. (B) Liposomes carrying different chemotherapies were formulated with unique barcodes. After administering the liposomes in vivo, tissues were removed and cells were stained with propidium iodine (PI), which distinguishes dead cells from live cells. Barcodes enriched in dead cells corresponded to successful chemotherapy agents.
suggests that gene expression in cultured cells may not reliably predict gene expression in primary cells or in vivo [42].

Nanoparticle barcoding enables simultaneous analysis of >100 nanoparticles in vivo

The lines of evidence described above suggest screening nanoparticles directly in vivo would be useful. However, the expensive nature of in vivo experiments has limited the field to testing a few in vivo. This is a universal problem in nanomedicine, and it has driven groups to design systems that facilitate high throughput in vivo nanoparticle screens. In all cases, these systems utilize ‘multiplexed’ signals, which are signals that can be quantified without interfering with one another. In the simplest case, nanoparticle 1, with chemical structure 1, is ‘barcoded’ (i.e., ‘tagged’) with signal 1; nanoparticle N, with chemical structure N, is barcoded with signal N. The nanoparticles are co-administered, and later, signals 1 and N are quantified simultaneously. In one example, scientists isotopically-barcoded silver nanoparticles with different functional peptides to evaluate potential targeting ligands [43]. Another group has used quantum dots to barcode polymeric nanoparticles; in vivo screens were performed to understand how nanoparticle surface modifications altered delivery through the blood brain barrier [44].

One multiplexed signal that has been used by groups (including ours) is DNA [33,45,46]. DNA is an excellent molecule for multiplexing. The number of different barcodes scales exponentially faster than any other signaling molecule; 4N different barcodes can be generated with a DNA sequence N nucleotides long. An 8 nucleotide barcode can create 65,536 different sequences. Additionally, DNA sequence readouts are easy to analyze. The cost of DNA sequencing has decreased more rapidly than Moore’s Law; it is now possible to generate 400 million DNA reads for $2000 using a machine that fits on a desktop. Because DNA sequencing has become an increasingly more common tool in several fields [47], easy to use software packages are readily available to help analyze and interpret the data.

Two distinct nanoparticle DNA barcoding systems have been reported to date [33,45,46]. In one example, DNA was formulated into liposomes alongside different chemotherapies [45] (Figure 2B). By varying the primers or altering the barcode length, different sequences were detected using gel electrophoresis or real time PCR. The authors tested several drugs at once in vivo and concluded that the DNA barcodes found in dead cells corresponded to nanoparticles containing effective chemotherapies. This method may be used in the future to test hundreds of different cancer drugs in vivo. We developed a separate DNA barcoding system that utilizes next generation sequencing. Our barcodes contain (i) universal primer sites, (ii) a 7 nucleotide region with fully randomized sequences (to monitor for biased PCR amplification), and (iii) an 8 nucleotide barcode region in the center [46] (Figure 3A). Initially, we demonstrated that this system predicted siRNA delivery in vivo and generated DNA sequencing outputs that were linear with respect to the administered DNA [46]. Later, we demonstrated that this system, which we named Joint Rapid DNA Analysis of Nanoparticles (JORDAN), could simultaneously analyze over 150 nanoparticles simultaneously in vivo [33] (Figure 3B).

Although there are different ways to design DNA barcodes, specific traits help increase the robustness of the data. Most importantly, universal primer sites — which are primer sites that do not change - confer an important advantage; maximizing the chance all barcodes are amplified in an unbiased way. This seems counterintuitive, but the vast majority of a DNA barcode should be identical; the barcode region of the DNA should be small. Adding chemical modifications including phosphorothioates to the 5’ and 3’ termini increase barcode stability. The barcode regions should also be designed to work together on solid phase next generation sequencing (NGS) platforms like Illumina. Since solid phase NGS relies on fluorophores (Figure 3C), each individual barcode must have a ‘base distance’ of at least 3; in other words, each barcode must be different from all other barcodes at 3 of the 8 positions (Figure 3D). We have designed >200 barcodes with base distances of 3 or more [33]. It is also critical to sequence the DNA input administered to the animals. This allows us to normalize the data and compare different cell types to each other within an experiment. Finally, like all big data systems, nanoparticle DNA barcoding experiments should include proper controls in each experiment. Two examples include a liposome loaded with caffeine (instead of a chemotherapeutic) [45] and a naked barcode [33]; in both cases, the negative controls performed poorly relative to the experimental conditions, as expected. A detailed protocol describing barcode design and a nanoparticle bioinformatics pipeline is available on DahlmanLab.org.

DNA barcodes enable nanotechnologists to track how hundreds of distinct nanoparticles deliver drugs in vivo for the first time; this enables nanoparticle studies that were not feasible using traditional methods. As an example, we compared how dozens of different nanoparticles delivered DNA to 8 different cell sub-types in the spleen. Using unbiased Euclidean clustering — a common bioinformatics technique that analyzes large datasets — we found that specific types of immune cells tended to be targeted by the same nanoparticles [33].

DNA barcodes have enabled large scale experiments in fields as varied as oncology, developmental biology, and
viral delivery. As an example, by using the single guide RNA (sgRNA) sequence as a “barcode” to denote a gene that was knocked out, scientists performed whole genome screens to identify genes and non-coding regions that regulate cellular response to cancer and immunotherapies [48–51]. More specifically, a system named Genome-Scale CRISPR-Cas9 Knockout Screening (GeCKO) helped determine which genes promote resistance to anti-cancer drugs [50]. GeCKO was used to target 18,080 genes in the same experiment; multiple sgRNAs were designed per gene. The pool of sgRNAs were administered to human cancer cell lines that also expressed Cas9; as a result, each cell on average had no gene knocked out, or 1 gene knocked out. After adding an anti-cancer drug, cells were allowed to grow; enriched sgRNA sequences in live cells...
corresponded to genes that increased drug resistance (Figure 4A). The same approach has been used to study metastasis in vivo. In this case, a pool of tumor cells were edited using GeCKO and administered in the hind limb of a mouse. After waiting a period of time, the authors isolated metastatic tumor cells from different organs, and thereby identified genes that promoted metastasis [52]. Similar approaches have been used to identify genes that suppress primary tumor growth in the liver and brain (Figure 4B) [53,54]. Scientists also designed a combinatorial DNA barcode system named ‘PolyLox’ that identified stem and progenitor cells that led to the development of the immune system [55]. Finally, there are reports of ‘capsid shuffling’ and other combinatorial cloning strategies; these use sequences that code for amino acids as barcodes that denote which amino acids successfully enabled Adeno-associated viral (AAV) vectors that facilitate cellular entry and DNA delivery [56–58]. Given that DNA is a highly efficient method to store and access information, it is likely that many other biology and engineering fields will use DNA barcodes in the future.

**Outlook**

RNA therapies have tremendous clinical potential but will require on-target drug delivery to reach the clinic. siRNA drugs in the liver are a great example; several patient populations are already benefiting from treatments that silence genes in hepatocytes. However, without new delivery vehicles, the clinical impact will be limited to local injection, and to the liver [19]. To target new cell types, it is important to test as many nanoparticles as possible. Traditional screening methods test delivery vehicles in vitro; however, in vitro particle screens may inaccurately predict delivery in vivo. In conjunction with rapid nanoparticle synthesis, high-throughput in vivo nanoparticle screening methods will allow scientists to track more particles simultaneously than ever before. Over time, this may allow scientists to better understand how nanoparticles behave in the body. We envision a time where enough nanoparticles have been tested in vivo that nanoparticles which target a given cell type can be rationally designed. We find it likely these high throughput approaches will help accelerate the development of new genetic therapies.

**Funding**

M.L., C.D.S. and J.E.D. were funded by Georgia Tech startup funds (awarded to J.E.D.). C.D.S. was funded by the NIH-sponsored Research Training Program in Immunoengineering (T32EB021962). This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Research 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.).

Acknowledgments
The authors thank Nirav N. Shaw for his helpful input.

Author contributions
All authors wrote and reviewed the paper.

Conflict of interest statement
There are no competing interests to declare.

References
Papers of particular interest, published within the period of review, have been highlighted as:
* of special interest
** of outstanding interest


An early demonstration that lipid nanoparticles can safely deliver siRNAs that treat disease in human patients. These data were later confirmed in Phase III trials, which are reviewed in Ref. [9].


This paper describes the rational improvement of siRNA sequences. Using biochemistry, the authors design siRNAs that confer long-lasting silencing in vivo. It is possible that these sequence modifications will enable once-yearly injections of siRNA to treat disease.


First nanoparticle DNA barcoding study published. This – along with reference 46 – demonstrate that DNA barcoding can be used to track many nanoparticles simultaneously in vivo.


Important review that outlines advances in next generation sequencing.


This paper describes GeCKO, a system now used throughout biology to simultaneously study how every gene in the genome affects a phenotype.


