

Modern Drug Discovery: Why is the drug development pipeline full of expensive failures?

sitn.hms.harvard.edu/flash/2020/modern-drug-discovery-why-is-the-drug-development-pipeline-full-of-expensive-failures/

April 21, 2020

by Sebastian Rowe

What would you do if I gave you \$2.6 billion? For pharmaceutical companies, the answer is develop a drug. Not a handful, not a few... a single drug. \$2.6 billion is the estimated development cost of each Food and Drug Administration (FDA)-approved drug. For perspective, 48 new medicines were FDA approved in 2019. Therefore, the total development cost of all the approved medicines amounted to \$125 billion; a figure roughly equivalent to the 2019 GDP of Nebraska!

The process of discovering a drug involves many scientists running thousands of experiments over many years (Figure 1). However, technological advances in various steps of this process have not translated into a dramatic increase in FDA-approved drugs. To understand why so many reported breakthroughs fizzle out and why the cost of drug discovery remains so high, let's walk through these steps to tackle an Unknown Human Disease (UHD).



Figure 1: The process of discovery, from finding a target to FDA approval. Each step in this process normally takes years and collaboration across many scientific fields.

Finding a target

How, and where, might you begin the search for a drug to help with the UHD? In living beings, there are many kinds of molecules, but for the most part, they can be classified into four groups: DNA, RNA, proteins, and metabolites (Figure 2). In the modern age, one will likely use an “omics” approach to start the search. “Omics” are a group of techniques that use large data sets to identify patterns in one of the four groups of molecules, but the starting point for most target searches is genomics (DNA).

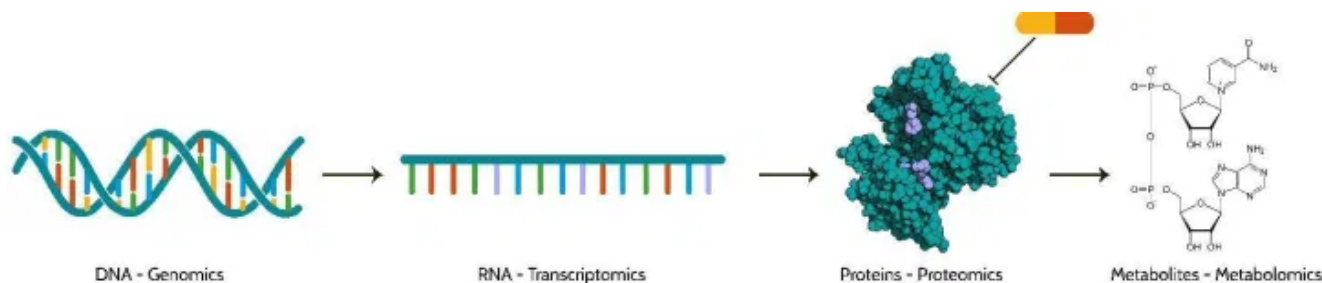


Figure 2: The molecules of our body. A gene's DNA is used to make a copy of RNA. This RNA is used as the blueprint to make copies of a protein encoded in the gene. Proteins do many different things in the body, but one of the most common roles is to use and create metabolites such as vitamins and sugars.

Genomics concerns itself with identifying what genes are associated with a specific disease. Hundreds of thousands of sample human genomes have been deposited into databases known as biobanks. These genomes are tagged with the medical history of volunteers, and scientists can analyze the data to discover what genes are associated with the UHD. Despite the great promise genomics has in helping us better understand disease, genes are not good drug targets.

Luckily, each gene usually codes for only a small set of proteins, and proteins are good drug targets. To borrow an explanation from former Harvard Professor and current President of the Novartis Institutes for BioMedical Research (NIBR) Jay Bradner, M.D., proteins are like pasta. They are better with certain fillings and they come in many, many shapes. A good drug, meanwhile, is like a good filling—it fits into the pasta well, it changes the flavor of the pasta, as well as the experience of the person eating it (the patient) for the better. Some pastas are easier to fill than others; the same is true for proteins. Currently, only 10% of human proteins are known to be “druggable,” or likely to be a good drug target (Figure 3). Because not every protein is druggable, breakthroughs in discovering the cause of the UHD do not always provide a good basis for drug discovery.

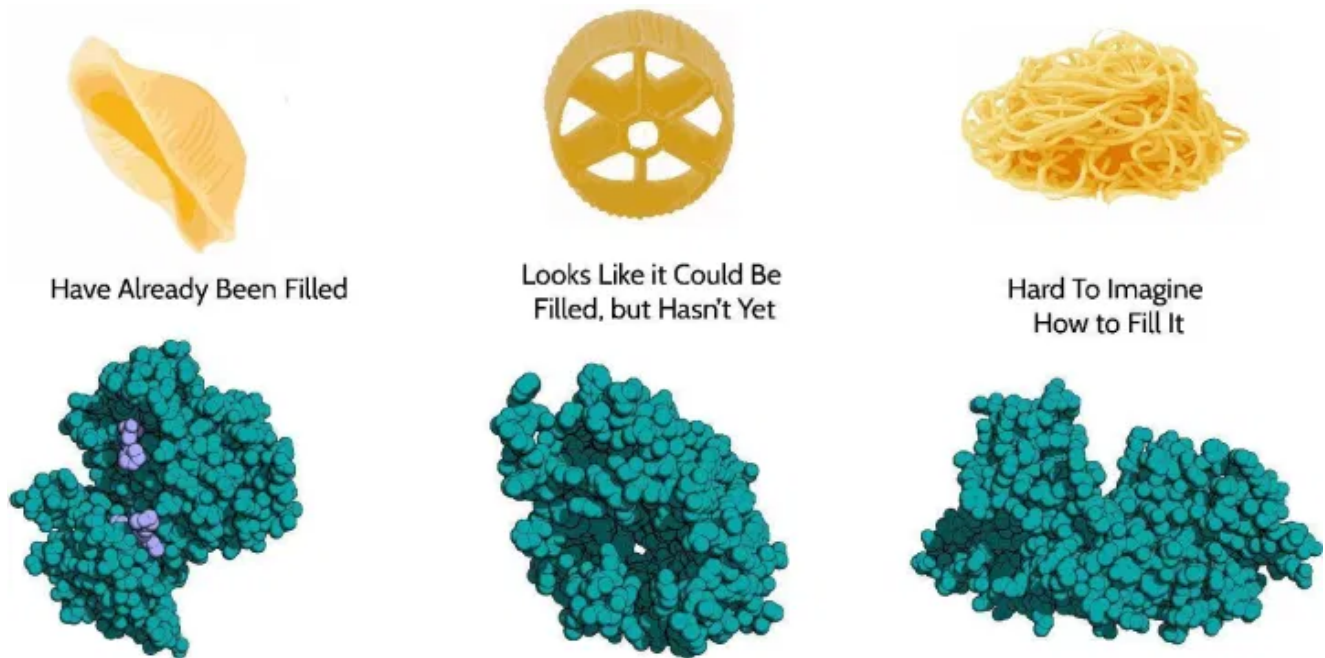


Figure 3: Different kinds of proteins (pastas) and their ease of filling. On the left is a protein kinase named EGFR (in teal) that is mutated in cancers and a drug called gefitinib (in light purple) that is used to treat breast and lung cancer. The middle is the protein BamA that is essential in certain bacteria and is a potential antibiotic target. On the right is the human protein TBX3 that, when mutated, can cause developmental defects and is associated with cancer metastasis.

How to find a drug?

Let's say scientists have found the gene that codes for a pasta protein (Your Favorite Protein – YFP) that they think is involved in the UHD and, importantly, that it can be targeted with a drug. How does one find a drug that will interact with it? If there isn't already a chemical known to interact with YFP, one can decide to test hundreds of thousands of chemicals quickly using robots in a process called High Throughput Screening (HTS). To decide what chemicals to test, medicinal chemists often use a computational technique called virtual screening that uses data from past screens to predict what molecules will interact. The library of chemical compounds is then screened for interactions with YFP; any compounds which interact are known as “hits” (Figure 4).

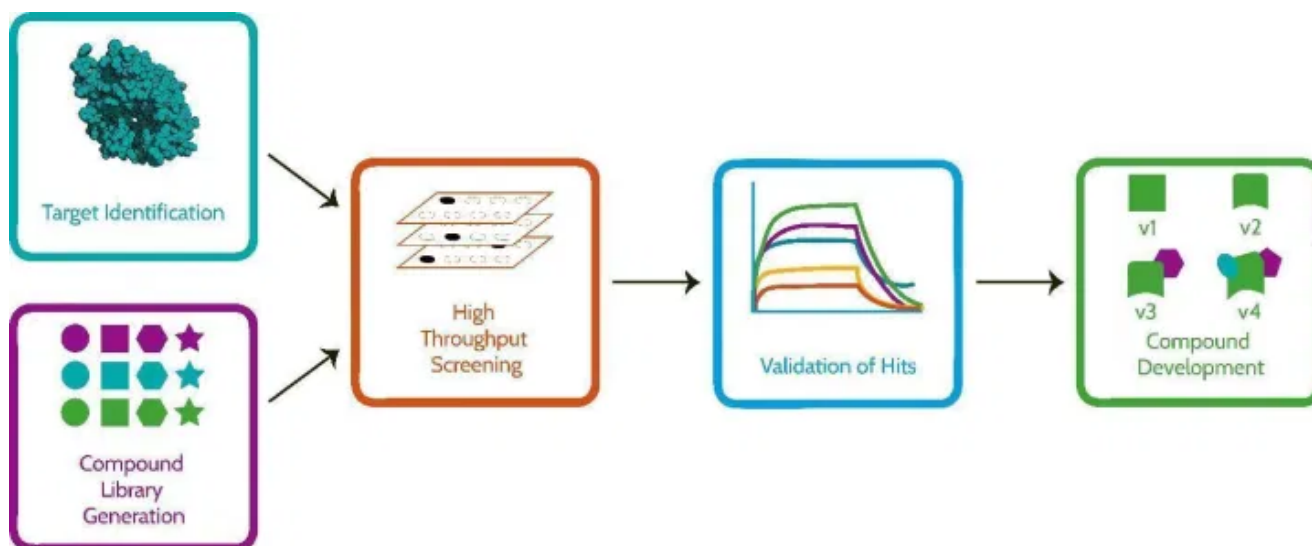


Figure 4: The pipeline of HTS. In the first step of HTS, one must identify and generate the target protein and a large library of chemical compounds. Then one must have a screen that gives measurable output on if a chemical interacts with the target. This screen is run many times over on all the compounds using robots. Any compound that is a hit in this screen is then validated and optimized.

Validation of hits can be carried out in a variety of ways, but at its core, hit confirmation depends on using a variety of biochemical assays to understand the interaction between the hit and YFP. Most of the hits will unfortunately interact in ways that are not relevant to the UHD, such as filling the wrong part of the protein. Any hits which survive validation are considered lead compounds. At this point, one has spent millions of dollars and years of research, yet most of the hits will fail to turn into a drug.

Turning a hit into a drug

Lead compounds from HTS are like an unseasoned, unprepared filling for pasta. They may work somewhat well but are often more useful as the foundation of a tastier filling. Chemists have developed many tricks to create drugs that are more effective. Oftentimes, these tricks involve finding the right combination of modifications to allow the drug to be more specific for its target. And nearly 100% of the time, the drug needs to be optimized to overcome many “logistical” challenges such as surviving the digestive tract and being soluble in the bloodstream. Furthermore, one must prove that the compound will cure the disease or alleviate symptoms in human cells and animal models (Figure 5).

The next step of drug discovery is often euphemistically referred to as the “valley of death”. The “valley” is the gap between drugs that work in Petri dishes and mice and drugs that actually work in the human body. The main obstacle for candidate drugs is toxicity in clinical trials (Figure 5). Clinical trials, which occur in three separate, sequential phases, take years to finish and on average only 1 in 20 drugs will perform well enough to even warrant entry into another complex set of processes for FDA approval. Therefore, the best bet to find a treatment for a UHD isn’t to create one optimized potential drug but 10 to 20. Here is where

the biggest portion of the \$2.6 billion price tag comes from: the litany of lead compounds that failed to validate, could not be optimized, floundered in human cells and animal models, or flopped in clinical trials.



Figure 5: Safety and efficacy testing prevent toxic compounds from being approved. Lead compounds are tested for efficacy in cell cultures and animal models. Next come Phase I trials where the candidate drug is tested for toxicity in healthy volunteers. In Phase II a small group of people with the UHD are treated with the compound. Successful results from Phase II must be repeated over a larger population in Phase III trials.

Takeaways and the future of drug discovery

Drug discovery is expensive, time consuming, and does not guarantee a novel drug for treating disease. But drug discovery is vitally important. Reassuringly, the number of new FDA-approved drugs has been increasing, and the current 5-year annual approval rate average is double what it was in 2009. New advances in modeling the human body such as stem cell technology and organoids could prevent toxic drugs from reaching clinical trials and thus avoid costly failures. Improvements in virtual screening could greatly reduce the chemicals needed to be screened for finding lead compounds. New ways to effectively target DNA and RNA could help cure diseases without a good protein target. An exciting technology for targeting DNA, CRISPR, is a set of proteins that can be programmed to modify specific DNA. Recently, therapies that interfere with RNA function were approved by the FDA. Furthermore recent advancements in chemical biology may greatly increase the number of druggable proteins. We may not have a drug for every disease yet, but with time (and a lot of research!) we will find better drugs for more diseases.

Sebastian Rowe is a first-year Ph.D. student in the Chemical Biology Program at Harvard University.

Cover image: “68/365 — 3/08/12 -” by glaukos is licensed under [CC BY-NC 2.0](#)

For More Information:

- Companion techniques to genomics include transcriptomics (RNA), proteomics (proteins) and metabolomics (metabolites).
- There are important issues with current genomic studies.
- The thalidomide debacle shows the need for good clinical trials.
- Failed drug candidates (like thalidomide) can still be useful.

- The story of Nobel Laureate Tu Youyou is a great example of drug discovery in the 20th century.