ELSEVIER

R&D Solutions for PHARMA & LIFE SCIENCES

DRUG DISCOVERY & DEVELOPMENT

Drug Attrition in Check: Shifting Information Input to Where it Matters

Summary

Expenditures in pharmaceutical R&D are higher than ever, and attrition rates have reached 80-90%. The few compounds that make it to market have to carry the cost of the inefficient development process, but hope lies in improving the data that flows into lead discovery and optimization. How does the pharmaceutical industry need to adapt their information frameworks to ensure that the relevant data is accessible to the right researchers at the right time?

ELSEVIER

Since the early 1990s, R&D expenditure by U.S. pharmaceutical companies has grown exponentially.

High attrition, high costs

Over the last three decades, pharmaceutical output has not changed much (Figure 1). The number of companies involved in drug development has grown—many come and gone. Yet, the number of new chemical entities approved by the U.S. Food and Drug Administration has remained flat, even after the adoption of new technologies, such as recombinant DNA technologies, mining of DNA sequence libraries, and sequencing of the human genome (1). What has changed is the cost of making drugs. Since the early 1990s, annual R&D expenditure by U.S. pharmaceutical companies has grown exponentially, leveling off at over 50 billion US dollars (Figure 1). With attrition rates as high as 80–90% (2, 3), current drug discovery approaches do come up with new entities, but too many of these advance to later development stages without proper vetting. Most drug candidates fail, and those that are approved carry the cost of this inefficiency.

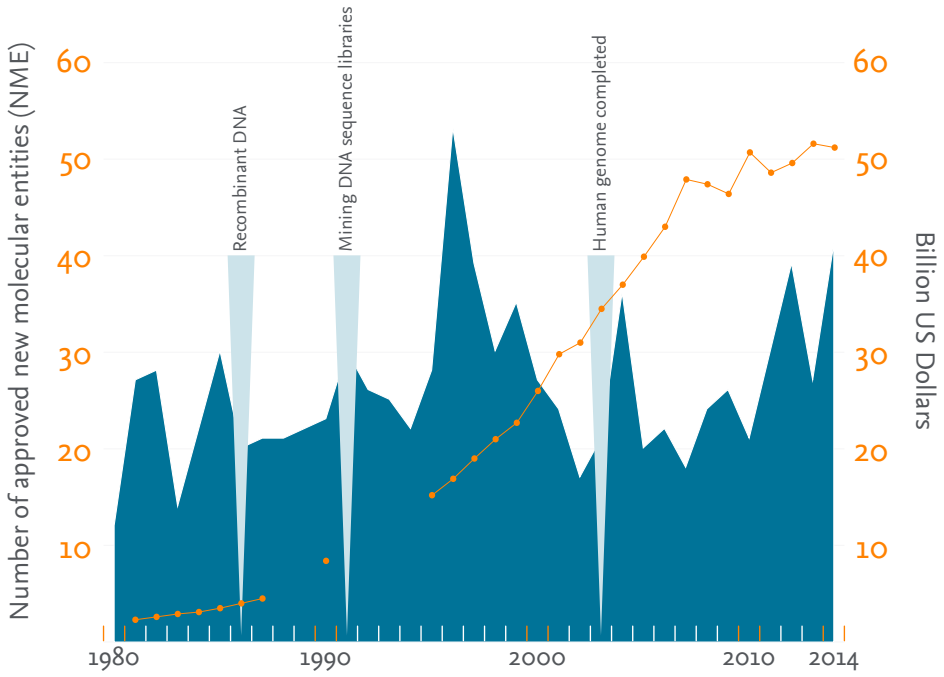


Figure 1. Number of new molecular entities (and new biological entities starting in 2004) approved by the U.S. FDA each year from 1980 to 2014 has remained relatively flat. Superimposed on the graph is the annual R&D expenditure reported by US PhRMA company members (orange line) and time points when technologies supporting target-oriented approaches to drug development became available (arrows). The outlier peak in NME approvals observed in 1996 stem from the review of backlogged FDA submissions after an additional 600 new drug reviewers and support staff were hired, funded by the Prescription Drug User Fee Act. Data extracted from the U.S. FDA website, DiMasi et al. 1991, and Statista (14–16).

First-in-class drugs are therapeutic compounds that are truly novel to the market; for example, a drug that uses a unique mechanism of action to exert its therapeutic effect. Such drugs grant the developer initial exclusivity in the market but carry the risk that the drug will not work in humans, will not be better than existing therapies, or has adverse effects that can only be uncovered with time.

Best-in-class drugs are **"follower"** drugs that build on a therapy that is already in the clinic and has been proven to work in patients, with the aim to deliver a better therapeutic effect than the original drug. The risk in developing these drugs is not that they may not work in patients but that they may not be an improvement.

Every step of the drug discovery process has hurdles and entails risks. Exacerbating these risks is the disconnect in scope between how drugs are developed and how they are used. A drug is scrutinized in a handful of interactions with wanted and unwanted targets, but then used in an organism where it encounters countless other interactions. Researchers are working to narrow this disconnect by informing drug discovery, design and optimization with everything known about the various interactions that a compound could have in exerting its effect on networks of biological molecules in a physiological environment. Injecting this knowledge into early drug development means investing in information: to build the context of a disease; to view data within that context; and to recognize failure early on. This way, only high-quality leads advance to the next stage, increasing chances of approval success and reducing costs.

Too narrow a focus

To date, therapeutic compounds continue to emerge from target-oriented development approaches that isolate lead discovery, design and optimization from the plethora of molecular interactions in a biological system. This is problematic for two reasons. First, simplifying the potential touchpoints of a drug makes development blind to the dynamics arising from the drug interacting with other biological molecules, or from biological molecules interacting with one another, or from the internal microenvironment influencing interactions – all of which can have repercussions for the efficacy of the drug.

Second, target-focused approaches have the effect of curbing pharmaceutical innovation. A study reported by Forbes (4) highlights that over 20% of about 1000 active oncology drug programs in 2012 concentrated on the same 8 targets: mTOR, c-MET, VEGF, c-Kit, PDGF, PI3K, HER2 and EGFR. The reason may be related to the difficulty of target validation, which is required in target-oriented approaches (1). Focusing on well-known targets to develop next-generation or "best in class" drugs is a simpler and preferred strategy over exploring new, uncharacterized targets. In fact, over the last decade, most "first in class" small molecule drugs emerged from compound screening strategies that were target-agnostic, whereas "follower" drugs tended to come from target-oriented drug discovery (4).

Target-oriented approaches have certainly not been fruitless. Examination of single drug–target pairs has fleshed out a detailed understanding of how compounds interact with biological molecules and has generated a list of structural features that impact both the binding of a compound to a target and the way a compound behaves in a physiological milieu. Nevertheless, this focus may narrow not only the full scope of interactions that can seriously impact the effectiveness of a drug candidate, it may also prevent developers from seeing the full scope of alternate ways to target disease.

A trend toward an expanded exploration scope has emerged with two relatively young disciplines, systems biology and systems chemistry. Applied to drug discovery and development, systems-based approaches use large amounts of high-quality data to predict interactions among biological molecules and compounds that may be relevant for the development of new therapies.

A disconnect of scope

Every individual cell has thousands of networks of interacting molecules regulating cellular metabolism, growth, reproduction and other critical survival functions. Normal cell phenotype depends on the meticulous orchestration of these molecular networks, which help the cell respond to its immediate environment. The sum of the phenotypes of all cells in a body determines the internal physiological state of that person: exact chemistry of the blood, function of tissues, coordinated activity of organs, etc. Imbalances at any level can lead to disease.

Drug development often operates under the assumption that a compound that decommissions one molecule

Evaluating information about the larger picture during drug candidate vetting could prevent late-stage failures.

in a molecular network causing an aberrant phenotype can taper or eliminate a disease. Thus, target-oriented development programs look for compounds that bind to a selected target molecule and work on altering the structures of a few top choices to improve their affinity and their efficacy, using assays developed to specifically measure these characteristics.

For example, recombinant biology techniques enable producing a target in sufficient quantities to perform standardized assays measuring how a compound alters the target's activity. Also, the target's three-dimensional structure can be constructed and used to conduct in silico binding simulations. Through multiple iterations of the drug design cycle—test a candidate compound, make changes to the structure to improve its affinity and/or efficacy, test it again—a drug candidate emerges, which is then assessed in preclinical models, such as cell cultures or animals, and then in clinical trials on humans.

This latter step, preclinical and clinical trials, represents the disconnect of scope between the generation of a lead and its testing in organisms. The first operates at the level of a single molecular interaction; the latter tests the drug candidate in the context of a full-fledged organism with all the complexity of a biological system. With an incomplete picture of the complex mechanisms that lead to disease, it is not uncommon that a drug fails at this stage. Unfortunately, preclinical and clinical trials are also the most expensive aspect of drug development. Failure here means failure of a much greater investment than if problems with a lead are recognized at earlier stages.

It stands to reason that evaluating information about the larger picture during the actual vetting of a drug candidate could prevent late-stage failures. The vetting should include questions like does the isolated interaction between lead compound and target behave the same within an interlinked network of biological molecules? What environmental conditions will hinder the drug on its path from route of administration to target? Zooming out from the single drug–target interaction to include knowledge about influential molecular networks and parameters at cellular, tissue, and even organism level can support new models that provide answers to these questions.

There are still gaps in the basic understanding of which and how molecules interact within a given network, cell, or tissue. Nevertheless, data are accumulating that can elucidate a broader landscape of molecular interactions and describe how those interactions contribute to health or disease. As knowledge grows, zooming out from the single compound–target pair to complement target-oriented drug development with an expanded exploration of complete systems can lead to the discovery of new therapies and support progression of better drug candidates through the development process.

Expanding perspectives

The exploratory space of medicinal chemistry encompasses the interaction between a biological landscape containing all druggable targets relevant to a therapeutic area, and a chemical landscape containing all compounds that modulate the behavior of those targets. Ideally, the complete collection of targets and compounds would be known, but in reality only portions of both landscapes are characterized, and even smaller subsets of compounds and targets are used in drug development.

Target promiscuity is the susceptibility of a biological molecule to bind and be affected by more than one compound.

Compound promiscuity is the ability of a compound to specifically interact with more than one biological molecule.

Functional correlations among targets refers to any interaction among biological molecules that influences the effect of a compound; e.g., networks that create functional redundancy, interactions with surrounding non-target molecules relevant to microenvironment.

Functional correlations among compounds are compound–compound interactions that impact their potential therapeutic effect.

Target-oriented approaches extract one target from the biological landscape and explore all the compounds that bind the target to identify and then optimize a lead with maximum specificity—the compound preferentially binds one target—and potency—the compound binds that target strongly and has a significant impact on target activity. This is a one-dimensional exploration of target promiscuity (highlighted portion of Figure 2). The practical advantage of this approach is that it pares down the exploratory space of a development program to a simple, linear and tractable link between disease and drug. This link can be further broken down to a set of well-understood design criteria and well-understood measures of drug impact that can be evaluated in a systematic and iterative drug design cycle.

However, there are still three other interaction dimensions of the exploratory space available to medicinal chemistry—compound promiscuity, functional correlations among targets, and functional correlations among compounds (Figure 2). Each dimension has the potential to reduce the disconnect between the development and use of a drug by better approximating the complexity of disease mechanisms and the internal physiological environment of a patient. In this way, potential problems with leads are highlighted and suboptimal leads can be removed from the pipeline early to contain costs.

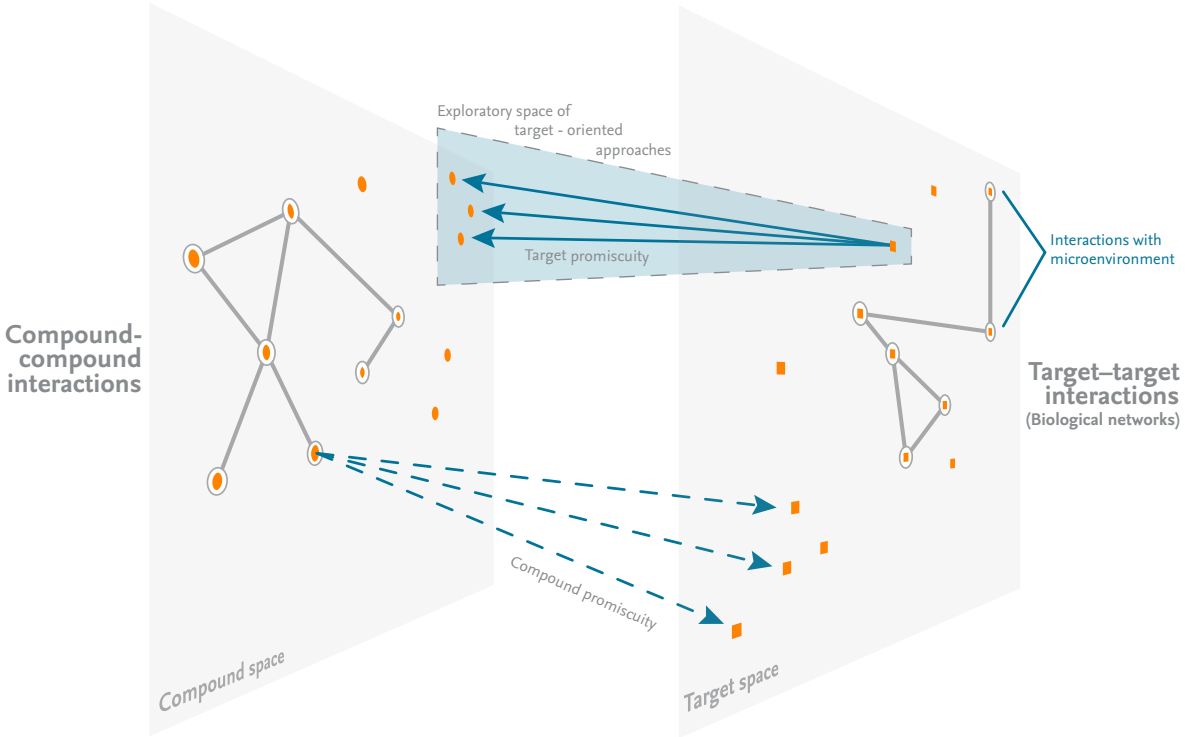
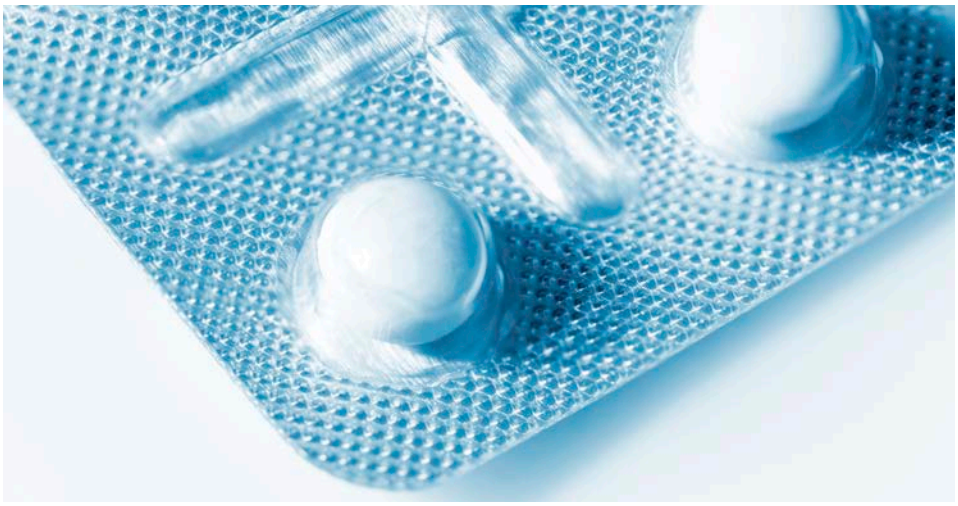


Figure 2. The exploratory space available to medicinal chemistry includes compounds and targets that interact in multiple dimensions. The most commonly explored dimension is target promiscuity: which compounds bind to a target? Equally valuable for exploration are the interaction dimensions encompassing compound promiscuity (which targets does a compound bind?), compound–compound interactions, and interactions among targets as well as other molecules present in the microenvironment of a target.



Each dimension also has the potential to reveal innovative therapies currently out of reach because target-oriented development points to a narrow subset of drug action mechanisms. Consider first compound promiscuity. Despite being developed to bind and modulate a single target, FDA-approved drugs have been shown to interact with an average of six molecular targets (5). Generally considered a negative attribute for a drug, information about promiscuous binding can be leveraged to develop much needed multi-target drugs that can curb resistance development. Functional correlations among biological molecules that alter the response of a target to a drug may also lead to new therapies. For example, molecular crosstalk between tumor cells and surrounding non-tumor cells contributes significantly to tumor growth, metastasis and resistance to therapy (6). This means that heterogeneity of cells within a tumor and in the environment immediately surrounding the tumor play a potentially critical role in drug efficacy. Understanding and modulating the mechanisms of this crosstalk could be a new therapeutic path. Finally, complementing the goal to create multi-target drugs are combinatorial therapies in which two or more compounds are combined to generate a polypharmacological effect that is more effective than hitting a single target. Creating such therapies requires examining interactions and functional correlations among combined compounds.

Considering the entire system

Systems biology and systems chemistry are developing tools and methods to leverage massive sets of 'omics data to explore the complete informational space relevant to medicinal chemistry. The goal is to capture the complexity of biological entities and their interactions with chemicals, and then capitalize on the properties that emerge from these interconnected systems. This enables a holistic view of an organism, tissue, cell or network of biological molecules responding to perturbations triggered by one or more chemicals that is much more than the sum of its components (6). Both are data-driven sciences with methodologies to summarize, visualize and interpret data across multiple scales (e.g., molecules, networks of molecules, cells, immediate surroundings of cells, tissues, an organism) and concepts (e.g., proteomics, genomics, epigenomics, metabolomics). Critical to the predictive power of systems biology and chemistry is the use of large amounts of high-quality data and an understanding of the context in which these data are produced and interpreted.

Systems-based approaches can complement conventional target-oriented approaches to decipher the complexity of molecular interactions within the context of heterogeneous microenvironments. In doing so, these methodologies offer a novel perspective that can lead to innovative therapeutic mechanisms, more sophisticated characterization of patients to better match them to therapies and combination therapies or means to bypass drug resistance (6).

Identifying compounds that modulate disease

Phenotypic screening of compounds, a standard method in pharmaceutical research before the 1980s when recombinant DNA techniques were introduced, has regained popularity in drug lead discovery. Taking a step back from the molecular theatre of drug–target interactions, phenotypic drug discovery tests compounds on cell cultures and small organisms (e.g., worms, flies,

zebrafish) to identify hits that generate a desired biological effect. This biological effect may be measured as changes in the physical properties of cells or the localization of its components, changes in the production of RNA, proteins, and metabolites, or any other measure that can be linked to the disease of interest (3).

Unlike target-directed drug discovery, targets are not known and readouts measure the result of multiple targets and pathways that are simultaneous interrogated. An advantage of phenotypic drug discovery is that it assesses the capacity of one or more compounds to modulate disease rather than just the ability to bind a target. This is a reason for the renewed interest in phenotypic screening as it bridges the disconnect between development and use of a drug already at the point of lead discovery. Phenotypic screening has a greater potential to lead to novel targets, increase chemical diversity of compounds being evaluated as drug leads, and uncover novel mechanisms of action (7).

Phenotypic screening is particularly promising for assessing the microenvironment and its impact on drug action. Co-cultures of different cell types are used to evaluate if the presence of cells surrounding those to be treated with a compound impacts responsiveness. For example, it has been shown that the anti-tumor effect of gefitinib (Iressa, AstraZeneca and Teva), a targeted lung cancer medicine, is attenuated by the presence of fibroblasts co-cultured with lung cancer cells (8). In fact, culturing tumor cells with stromal cells, such as fibroblasts, can cause them to arrange differently on a two-dimensional plane and in a three-dimensional matrix, which can impact their uptake of a given compound (7). Furthermore, a systematic evaluation of co-cultures using different stromal and cancer cell types demonstrated that stromal cells commonly mediate resistance to therapeutic agents (9).

Informing lead optimization with large amounts of data

Data from a phenotypic screen must be translated into guidance for optimizing the leads identified in the screen. Typically, a variety of experimental methods are used to identify the target or possible mechanism of action, but in silico approaches that take advantage of large databases can also predict potential targets or mechanisms of action. For example, databases can be mined to predict targets based on structural similarities shared between the identified lead compound and well-characterized compounds (10). Another intriguing approach has been to construct response profile databases for assay systems. That is, a system such as an optimized cell culture assay, is exposed to a large library of known bioactive compounds for which interactions with biological molecules and/or mechanisms of action are known. Detailed response profiles are recorded and made searchable so that, by comparing experimental profiles to the database, hypotheses about target and action can be generated (7).

Network- or systems-based approaches can support a more comprehensive conceptual framework for drug development but they rely heavily on large amounts of high-quality data that must be interpreted within the context of a meaningful knowledge base about the disease, pharmacokinetics and pharmacodynamics. Mechanistic models are informative only if constructed based on solid, empirically determined parameters, which must be extracted from the literature or defined. For example, Bianconi et al. (11) constructed a model to explore interactions between two pathways known to play a role in lung cancer. The model included 45 parameters, each of which required establishing values based on available literature. These mechanistic models were then tested experimentally, improved based on results, and then tested again.

Similarly data-intensive was the work by Liu et al. (12). They used over 1.5 Gb of raw data to construct and validate a module of correlated genes identified in expression data from 58 samples of lung cancer tumor tissue and adjacent healthy tissue via weighted correlation network analysis (WCNA). Six genes in the module proved to be predictive markers of the disease. Another example is the work of Yabuuchi et al. (13), who used over 15,000 pairs of kinases and kinase inhibitors to construct and validate a virtual screening model that assesses the likelihood of a successful novel compound–target interaction.

The value of investing in information

As the search for novel pharmacotherapies taps into this expanded exploratory space, the magnitude, diversity and sources of data needed to elucidate interactions in complex biological systems, understand their contribution to health or disease, and ultimately identify successful therapies will grow. Effective data-driven drug discovery, where meaningful knowledge informs early stages of drug development, will require the adoption of more comprehensive information frameworks. More than data repositories, these frameworks will integrate and enable transparent data generation or sourcing (where do these data come from?), efficient data discovery and management (is this all the information relevant to a question?), and the use of information across a range of scientific domains (what do these data mean and can they be used in this context?).

Pharmaceutical companies already house vast collections of in-house data. One function of these information frameworks will be to host a single repository for data generated internally and collected from third-party sources that are continually updated and accessible company-wide. This body of information, however, can support data-driven methodologies only if integrated with information that clarifies the fit of any dataset to an overall disease picture. That is, drug developers must operate at the intersect of data from disparate sources and understand how these data relate to one another and to published information. Thus, these information frameworks will also need to incorporate structures and processes for suitable information stewardship, such as curation, normalization and quality control of data.

A final component of these information frameworks will be implementation of an extensive knowledge base for data interpretation. Identifying synergies across datasets and across knowledge from different disciplines can only be facilitated by a framework that organizes highly granular data with unstructured information from public databases, published literature and patent content into a taxonomy of relationships among scientific concepts and terms. And this organizational taxonomy will need to adapt to ever-growing and changing scientific knowledge and terminology to guarantee that connections revealed remain relevant and accurate.

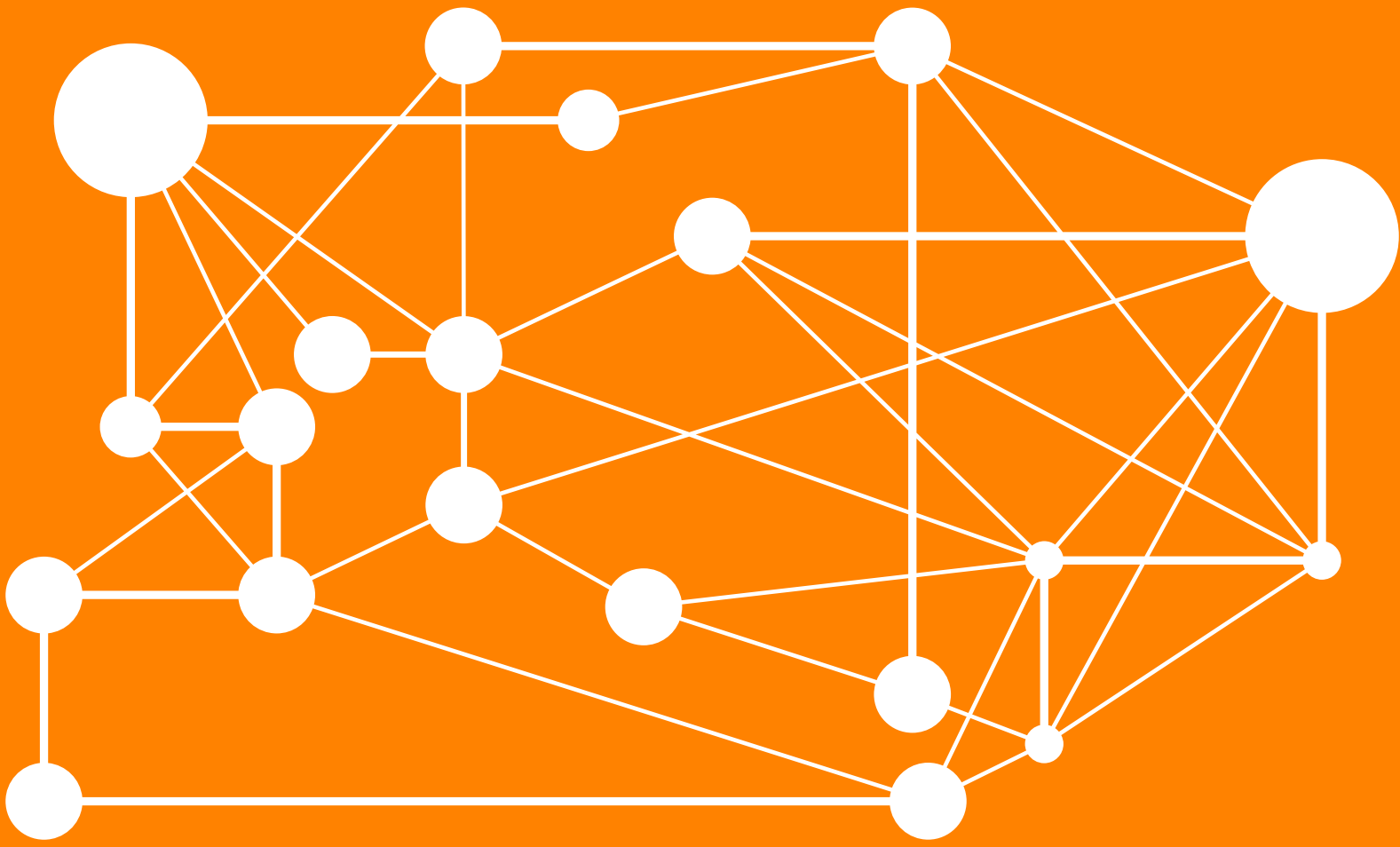
Such comprehensive information frameworks translate unstructured bits of knowledge into usable, structured data that can be combined and connected in a meaningful way. They allow scientists to make sense of the data they generate in a broader context and facilitate communication across the various disciplines that merge in a drug development program. And we cannot underestimate the value of empowering teams to work towards a common goal. In this way, these information frameworks provide a solid foundation to funnel as much information as is available into answering questions about the action of a compound in a complex biological system while it is being designed, thereby eliminating the risk of failure due to foreseeable issues. Investing in the expertise and resources to build such a framework is time-consuming and costly. Alternatively, it may be possible to curtail timelines and costs by leveraging collaborations with parties that already have the necessary expertise and structures to merge and organize scientific information. Regardless of the path, implementing an expanded scope for exploration and discovery of new strategies to tackle diseases is a critical step toward making data-driven drug discovery a reality.

Despite all the challenges, Dr. Herbert Köppen, a veteran of the pharmaceutical industry who has spearheaded the implementation of systems-based approaches for decades, sees this change in the way drugs are developed as inevitable: “Quite honestly, I see no way to escape this paradigm shift. The fact is, only drugs that provide a real therapeutic benefit will pay off research and development investment, and it is clear that the single-target approach no longer meets that demand.”

That is, pharmaceutical productivity can only be boosted if development efforts eliminate inefficiencies and generate enriched sets of properly vetted drug candidates to be tested in preclinical and clinical stages. In response to the question of what it will take for systems biology to be used routinely in drug development, he says: “It will take one research division head to have the foresight to make an investment along this line; it will take a lot of work to validate techniques and models; and it will take time for this field, which is in its infancy, to find strong footing.” However, exploring these unchartered waters may also usher in a new era in pharmaceutical innovation.

REFERENCES

1. Lee, J.A. and Berg, E.L. 2013. Neoclassic Drug Discovery: The Case for Lead Generation Using Phenotypic and Functional Approaches. J. Biomol. Screen. 18, 1143–1155.
2. Hay, M., Thomas, D.W., Craighead, J.L., Economides, C. and Rosenthal, J. 2014. Clinical Development Success Rates for Investigational Drugs. Nat. Biotechnol. 32, 40–51.
3. DiMasi, J.A., Grabowski, H.G. and Hansen, R.W. 2014. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. http://csdd.tufts.edu/news/complete_story/cost_study_press_event_webcast
4. Booth, B. 2012. Cancer Drug Targets: the March of the Lemmings. Forbes Mag. http://www.forbes.com/sites/brucebooth/2012/06/07/cancer-drug-targets-the-march-of-the-lemmings/#35f9930c3789>
5. Mestres, J., Gregori-Puigjane, E., Valverde, S. and Sole, R.V. 2009. The Topology of Drug–Target Interaction Networks: Implicit Dependence on Drug Properties and Target Families. Mol. Biosyst. 5, 1051–1057.
6. Werner, H.M.J., Mills, G.B. and Ram, P.T. 2014. Cancer Systems Biology: a Peek into the Future of Patient Care? Nat. Rev. Clin. Oncol. 11, 167–176.
7. Berg, E.L., Hsu, Y.-C. and Lee, J.A. 2014. Consideration of the Cellular Microenvironment: Physiologically Relevant Co-Culture Systems in Drug Discovery. Adv. Drug Deliv. Rev. 69, 190–204.
8. Yong, X. et al. 2014. Fibroblasts Weaken the Anti-Tumor Effect of Gefitinib on Co-Cultured Non-Small Cell Lung Cancer Cells. Chin. Med. J. (Engl). 127, 2091–2095.
9. Straussman, R. et al. 2012. Tumour Micro-Environment Elicits Innate Resistance to RAF Inhibitors through HGF Secretion. Nature 487, 500–504.
10. Lee, J. and Bogyo, M. 2013 Target Deconvolution Techniques in Modern Phenotypic Profiling. Curr. Opin. Chem. Biol. 17, 118–126.
11. Bianconi, F. et al. 2012. Computational Model of EGFR and IGF1R Pathways in Lung Cancer: a Systems Biology Approach for Translational Oncology. Biotechnol. Adv. 30, 142–153.
12. Liu, R., Cheng, Y., Yu, J., Lv, Q.-L. and Zhou, H.-H. 2015. Identification and Validation of Gene Module Associated with Lung Cancer through Coexpression Network Analysis. Gene 563, 56–62.
13. Yabuuchi, H. et al. 2011. Analysis of Multiple Compound–Protein Interactions Reveals Novel Bioactive Molecules. Mol. Syst. Biol. 7, 1–12.
14. U.S. Food and Drug Administration. Summary of NDA Approvals & Receipts, 1938 to the present. http://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/SummaryofNDAApprovalsReceipts1938tothepresent/default.htm
15. DiMasi, J.A., Hansen, R.W., Grabowski, H.G. and Lasagna, L. 1991. Cost of Innovation in the Pharmaceutical Industry. J. Health Econ. 10, 107–142.
16. Statista. Spending of the U.S. Pharmaceutical Industry on Research and Development at Home and Abroad from 1990 to 2014 (in Million U.S. Dollars). http://www.statista.com/statistics/265090/us-pharmaceutical-industry-spending-on-research-and-development/



R&D Solutions for Pharma & Life Sciences

Elsevier's R&D Solutions is a portfolio of tools that integrate data, analytics and technology capabilities to help pharmaceutical & life sciences customers realize the potential of data-driven drug discovery and development.

LEARN MORE

To learn more about Elsevier's R&D Solutions, visit elsevier.com/rd-solutions/pharma-and-life-sciences or contact your nearest Elsevier office.

ASIA AND AUSTRALIA
Tel: + 65 6349 0222
Email: sginfo@elsevier.com

JAPAN
Tel: + 81 3 5561 5034
Email: jpinfo@elsevier.com

KOREA AND TAIWAN
Tel: +82 2 6714 3000
Email: krinfo.corp@elsevier.com

EUROPE, MIDDLE EAST AND AFRICA
Tel: +31 20 485 3767
Email: nlinfo@elsevier.com

NORTH AMERICA, CENTRAL AMERICA AND CANADA
Tel: +1 888 615 4500
Email: usinfo@elsevier.com

SOUTH AMERICA
Tel: +55 21 3970 9300
Email: brinfo@elsevier.com

February 2017