Drug discovery is a tough business scientifically, the traditional steps taking nearly a decade. In this article we are concerned only with the very beginning of the process, the identification of a target for a new drug. It is this step, the choice of a gene product of clinical relevance, that is the greatest impediment to expanding the pharmaceutical arsenal. In the United States, only 20–30 new chemical entities are approved as drugs each year, and the picture is much the same in Europe. Of these, only a quarter act on targets not already hit by an existing drug. Why has the deciphering of the 25,000 or so genes in the human genome not swollen the ranks of new targets?

The major reason, of course, is that targets are of value for drug discovery only if they can be convincingly related to disease. Such validation really takes not the one year shown on standard pipeline charts (Fig. 1), but decades. For example, the discovery of statins as agents that lower cholesterol levels in the bloodstream rested on investigations that began 40 years earlier with a study showing the relationship between cholesterol level and vascular disease. The recent discovery of Gleevec, which has revolutionized the therapy of chronic myeloid leukaemia, certain gastrointestinal tumours and other cancers, was based on work that began in the 1950s (ref. 2). Thus, it has been argued that well-validated targets, the low-hanging fruit as it were, are simply exhausted.

Validation is generally not a one-step process, but is rather an edifice built from studies in epidemiology and disease physiology, and from the results of research with animal models. The one exception is mendelian disease. In these disorders, such as cystic fibrosis and sickle-cell anaemia, the inheritance of a mutation in a single gene can be incontrovertibly linked to a physical characteristic, or phenotype — in this case the disease. Of the 6,000 or so illnesses with a mendelian pattern of inheritance, the gene responsible has been identified in approximately 1,200 (refs 5, 6). Historically, pharmaceutical companies have not concentrated on these diseases, in some cases because the affected protein is not tractable to pharmaceutical approaches, in others, perhaps, because the number of people affected is small. But the powerful role of a single gene in mendelian disease can provide insight into complex diseases where the same gene accounts for part of the phenotype. Statin therapy, for example, was initially directed to patients with a genetic predisposition to excessive levels of blood cholesterol. But after the drug’s efficacy and safety had been tested, the treatment was extended to a wider population of patients who had the same condition but due to many causes.

In seeking new targets, we first need to think about how we describe disease. Medical textbooks are organized by organ system, and diseases are classified by their pathology or physiology. If we hope to come to grips with the heterogeneity of common disease and its consequences for the choice of targets for drug discovery, clinical manifestations of disease must be causally related to a molecular definition. A simple concatenation of all the molecular changes in a disease, for example a compilation of profiles of which genes are being transcribed, and when and where, would be chaotic. A further step of integration and interpretation is needed. Ideally, this would provide a ‘grammar’ that would apply right through from the discovery of a drug target, to its testing in animal models and finally to the treatment of patients. In this context, a grammar means a set of rules, not for organizing words into sentences but for translating gene products into medicines. Just as the grammar we use in our speech is embedded in language, a grammar of drug discovery should be embedded in biological systems.

Molecular pathways

Intracellular molecular signalling pathways provide such a grammar for the genome. These pathways are triggered by extracellular molecules that bind to receptors in the cell membrane, thereby switching on relay systems inside the cell. The upshot is gene activation, or inactivation, that affects a cell’s behaviour — its ability to grow or differentiate, for example, or to undergo division or self-destruction. The number of distinct signalling pathways recognized in human cells depends on the definition. If classified by the type of cell-surface receptor involved, it could be as low as 16 (ref. 9). Using definitions based on cell type or gene-family member, and variants thereof, it could be as many as 200 (ref. 10). In any case, the number of signalling pathways is far fewer than the 25,000 or so human genes.

The elements of four now-canonical molecular pathways are shown in Figure 2 (overleaf): these lead from the receptors for proteins of the Wnt, Hedgehog (Hh), transforming growth factor-β (TGF-β) and insulin/insulin-like growth factor (IGF) families. Only the core components are depicted here in a much simplified manner. The important point is that these pathways are conserved throughout evolution — in the fly and mouse, in the worm and in mammals. A grammar of drug discovery should reflect this conservation.

To realize the potential of the genome for identifying candidate drugs we must move beyond individual genes and proteins. The signalling pathways in cells provide the right level for such analyses.
most of the animal kingdom, in invertebrates and vertebrates, and so they evidently control some of the basic cellular functions of life.

The power of molecular pathways to provide a grammar is evident from studies of embryonic development. As first demonstrated in the fruitfly Drosophila, the language of development is best phrased by the grammar of molecular pathways\textsuperscript{11,12}. At a molecular level, the characteristic units (sentences, if you will) of fly development are written in the form of pathways that determine, for example, the first anterior–posterior patterning decision, or the subsequent division of the body plan into segments or the generation of wings, eyes and other tissues. In zebrafish and mouse, similar unitary modules, often using the same molecular pathways, also define elements of organ systems and function in embryonic development\textsuperscript{13}.

Fundamental cellular processes that control growth and cell death in adults are similarly regulated by canonical molecular pathways\textsuperscript{11,14}. Such pathways can be represented graphically\textsuperscript{15} and measured quantitatively\textsuperscript{16}. The order and interactions of pathway components can be defined by genetics and by protein chemistry. All in all, they can provide a grammar applicable to drug discovery and medicine, from target validation through to the clinic.

The link between disease and signalling pathway is best validated for genetic disorders. Table 1 lists several diseases caused by mutations in the genes of components of the Wnt, Hh, TGF-β and insulin/IGF pathways. Perturbation of the essential processes driven by these pathways is the cause of many diseases, including diseases with complex causes, such as diabetes and heart disease.

Perhaps the best example known so far is the perturbation of a single pathway that regulates cell growth — the insulin/IGF–AKT pathway (Fig. 2) — in a host of apparently unrelated diseases\textsuperscript{17}. The pathway is activated in many cancers. For example, an increase in levels of phosphatidylinositol 3-kinase (PI3K) or mutations in its regulatory subunit cause cancers of the ovary and stomach, and activation of AKT or a decrease in PTEN function cause tumours of the prostate gland.

Other growth phenomena also depend on the insulin/IGF pathway. Examples are the proliferation of smooth muscle that leads to a narrowing of arteries (restenosis)\textsuperscript{18} following treatment to unblock them, and the increase in the numbers of certain immune cells — T cells — during transplant rejection. A key downstream node in this pathway in all these disorders is the protein mTOR, which regulates the translation of messenger RNA into protein. In clinical trials, inhibition of mTOR with derivatives of the macrolide rapamycin, an immunosuppressive drug, is now proving effective against cancer and restenosis as well as in transplant rejection.

The key prediction of such logic is that the drug target can be a ‘sensitive node’, not necessarily the protein that is specifically perturbed — which may be hard to define or not readily ‘druggable’ using available treatments. For example, predisposition to colon cancer is due to mutations in APC, a component of the Wnt pathway (Fig. 2), a protein that is technically difficult to target. But compounds that block downstream interactions in this pathway, those of β-catenin with its co-activators, have been discovered, and are able to suppress growth of tumours arising because of mutations in APC\textsuperscript{19,20}.

How can we discover the druggable nodes in other pathways? One approach is to select points of intervention, based on the current understanding of the pathways and their druggable components, and begin building cell-free assays to screen for small molecules that will alter the activity of these targets. A less biased approach is to combine genetics with chemical screens using potential drugs. The premise of such an approach is that the effect of a drug resembles that of the genetic perturbation of its target\textsuperscript{21–23}. For example, screens in yeast have shown that the growth-inhibitory effect of a compound often matches that of a mutation in the gene encoding its target\textsuperscript{22}. In the chemical screen, libraries of drug-like compounds are applied to cells, or even whole organisms, to alter pathway activity, and their effects are then compared with those of known mutations. The presumptive target can be confirmed biochemically. Thus, in principle, a compendium of drugs could be established that interrupt or stimulate every pathway of interest.

**Next steps**

‘In principle’ is the crucial phrase here, for our current understanding of molecular pathways is insufficient as a platform for effective pharmaceutical discovery. Several biotechnology companies have focused on the known elements of a few key pathways to target them with new medicines. But for the genome to be translated into medicines with any reliability and regularity, far more work needs to be done. Defining the role of pathways in complex diseases will undoubtedly take many years. A first step is to define the full complexity of these signalling networks at a molecular level — their ‘systems biology’ — including activity specific to a particular cell type, dynamic feedback mechanisms, extensive inter-pathway connectivity, the kinetics of signalling, and, of course, their state of activation in disease. The following issues will have to be addressed.

- Pathway outputs have different effects in different contexts. For example, the pathway activated by another extracellular signal, fibroblast growth factor (FGF), controls cell division in many tissues but in others it regulates cell differentiation, shape or survival\textsuperscript{24}. Sonic hedgehog, one of the related Hh signalling...
molecules, defines cell fates in the embryonic neural tube, but promotes cell division in the embryonic cerebellum. The outcome may even vary for a single tissue. For example, when it acts in breast epithelium, TGF-β normally helps to prevent the malignant transformation of the cells, in that it hinders cell growth and multiplication and promotes cell suicide. But if the epithelial cells are transformed as a result of a mutation upstream of SMAD, for example, TGF-β enhances blood-vessel formation and tumour-cell invasiveness, which encourage tumour growth and dispersal. Understanding the druggable elements or nodes in pathways that can trigger such altered biological states should define new targets for treating many different diseases.

Pathways are dynamic, with feedback mechanisms. For example, FGF stimulation of airway formation in Drosophila is terminated by the intracellular activation of the protein Sprouty, which restricts the process to a specific location. Similarly, Wnt signalling causes the up-regulation of inhibitors of the Wnt receptor Fzd, which then inhibit the pathway. In organ systems, changes in feedback, or in its phasing with regard to the stimulus, can have adverse effects. For example, abnormal breathing patterns result when feedback to the chemoreceptor concerned is out of phase with its output. It is likely that feedback abnormalities have adverse consequences in a cellular context as well. A major goal is to define the molecular components that are responsible for maintaining physiological balance, as these proteins are likely to be key regulators of dynamic disease.

Pathways intersect. Work in classical genetics has defined the phenotypic consequences when different sites on the same pathway are perturbed. Pathways also engage in extensive cross-talk, with each other, however, and these interactions change over time, for example after stimulation. But despite the complexity of these interactions, they are finite: they can be mapped and the key nodes in the intersections identified. As we come to understand this cross-talk and its role in disease, a logic for combination therapy should emerge in which two or more pathways can be targeted simultaneously.

Pathway kinetics vary between cells. The quantitative flow of signals through pathways depends on the levels of the specific components of a pathway. For example, quantitative modelling shows that the level of axin (Fig. 2) is what determines the amplitude and duration of signals through the Wnt pathway. To create cell-specific modulators of pathways we will probably need to define these limiting steps and target them.

**Clinical relevance**

One advantage of a pathway-based clinical taxonomy is that it affords a mechanistic foundation for disease description. Historically, diseases have been categorized by organ pathology. Today we are moving towards the pan-genomic assays of gene expression, protein levels and the ways in which proteins become modified after they have been produced from mRNA. The results of these assays become easier to interpret when several elements of a pathway are affected. Subtle changes in an individual protein may be below the threshold of detection, but may be amplified by sophisticated computational methods that incorporate several pathway components.

Another advantage of focusing on pathways is the ability to choose the most appropriate set of patients for initial tests of possible treatments. For many pathways, people with specific genetic defects have been described (Table 1), but, because of their rarity, the discovery of a treatment has often been relegated to the ‘orphan drug’ category. The development of therapies for such patients would not only serve as a medical need, but would often be readily extrapolated to a wider population. For example, Gorlin’s syndrome, which results from mutations in the protein Patched-1 (Ptc in Fig. 2; a negative regulator of Hh signalling), causes a brain cancer called medulloblastoma and a common form of skin cancer. The syndrome itself only affects between 1 in 50,000 and 1 in 150,000 people. But many other tumours, including non-small-cell lung cancer, gut-related tumours and prostate cancer, depend on Hh signalling, even though they do not seem to have mutations in pathway components.

The rationale for extrapolating from genetic to sporadic illness is that nature is conservative: this is a safe bet. And there would be immense immediate benefit in tackling the rare diseases in themselves. Patients with uncommon disorders are often neglected, their only hope being that medicines designed and marketed to treat common disorders might coincidentally be able to treat theirs. In contrast, in the logic of molecular pathways, such patients would be viewed as key intermediaries in the drug-discovery process, providing proof of concept. Along the undoubtedly long route to making molecular pathways a useful grammar for medicine, essential steps will involve successfully treating these well-defined but rare diseases. This approach could not only bring a new order to the genome, but could also have a salutary ‘side effect’ — a refocusing of drug-discovery research on neglected diseases.

Mark C. Fishman and Jeffery A. Porter are at the Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

**Table 1** | Fundamental pathways and a partial list of disease links.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Associated diseases</th>
<th>Affected gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedgehog</td>
<td>Gorlin’s syndrome</td>
<td>Ptc1</td>
</tr>
<tr>
<td></td>
<td>Basal cell carcinoma</td>
<td>Ptc1</td>
</tr>
<tr>
<td></td>
<td>Medulloblastoma</td>
<td>Ptc1</td>
</tr>
<tr>
<td></td>
<td>Glioblastoma</td>
<td>Gli1</td>
</tr>
<tr>
<td>Wnt</td>
<td>Vitreoretinopathy</td>
<td>Fzd4</td>
</tr>
<tr>
<td></td>
<td>Norrie’s disease</td>
<td>NDPI</td>
</tr>
<tr>
<td></td>
<td>Colon</td>
<td>APC</td>
</tr>
<tr>
<td></td>
<td>Osteoparosis-pseudoglioma/osteopetrosis</td>
<td>LRPS</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Fibrodyplasia ossificans</td>
<td>BMP4</td>
</tr>
<tr>
<td></td>
<td>Haemorrhagic telangiectasia</td>
<td>Aik1</td>
</tr>
<tr>
<td></td>
<td>Pulmonary hypertension</td>
<td>BMP2</td>
</tr>
<tr>
<td></td>
<td>Juvenile polyposis syndrome</td>
<td>SMAD4</td>
</tr>
<tr>
<td></td>
<td>Pancreatic cancer</td>
<td>SMAD4</td>
</tr>
<tr>
<td>Insulin/IGF</td>
<td>Cowden’s disease</td>
<td>PTEN</td>
</tr>
<tr>
<td></td>
<td>Tuberous sclerosis</td>
<td>TSC1</td>
</tr>
<tr>
<td></td>
<td>Multiple cancers</td>
<td>PTEN, PI3K, AKT</td>
</tr>
</tbody>
</table>

10. www.biocarta.com/genes/allPathways.asp