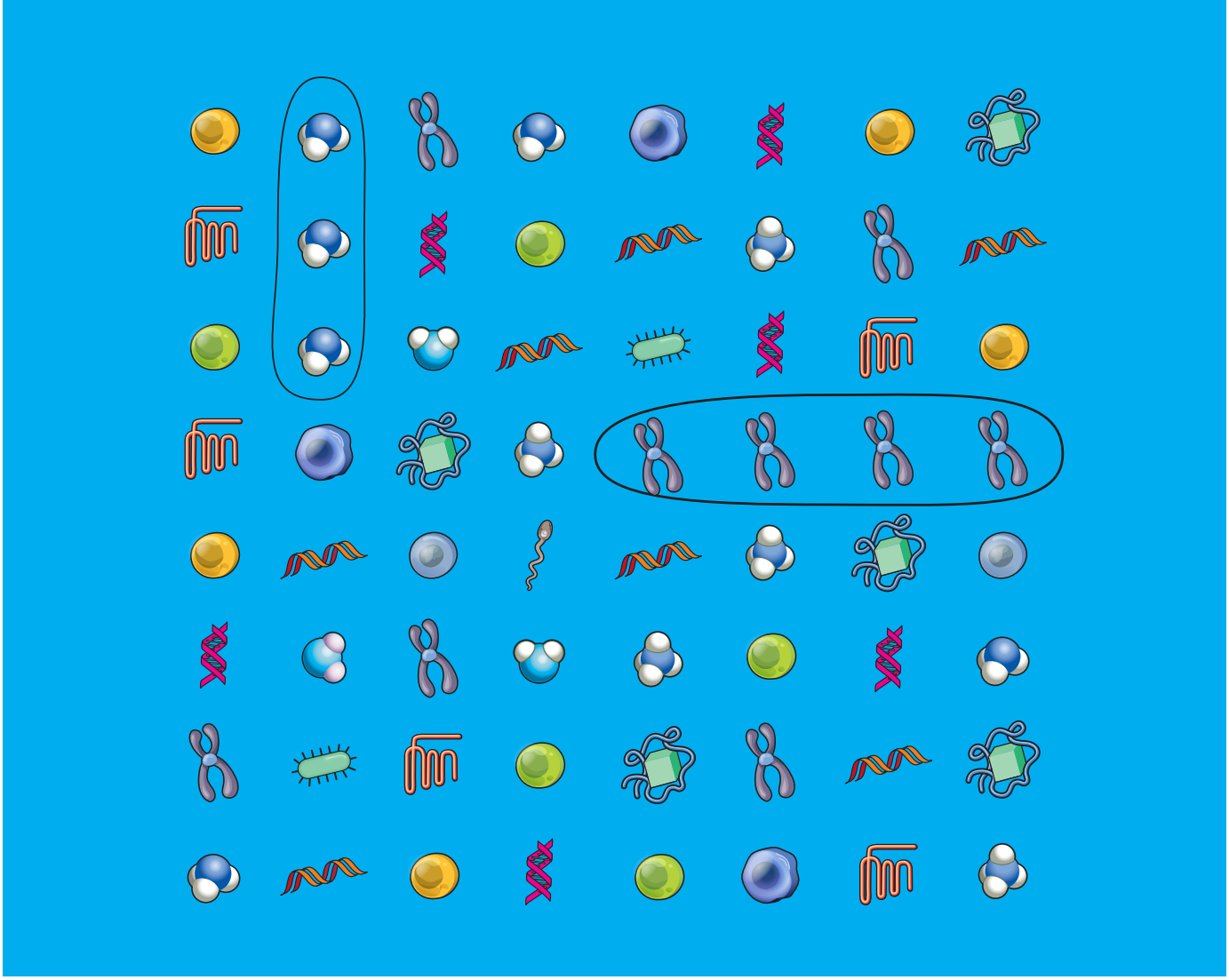


THE 'OMES PUZZLE

Where once there was the genome, now there are thousands of 'omes. *Nature* goes in search of the ones that matter.

BY MONYA BAKER

'Omics bashing is in fashion. In the past year, *The New York Times* and *The Wall Street Journal* have run pieces poking fun at the proliferation of scientific words ending in -ome, which now number in the thousands. One scientist has created a bad-omics generator, which randomly adds the suffix to a list of biological terms and generates eerily plausible titles for scientific papers (example: 'Sequencing the bacteriostaticome reveals insights into evolution and the environment'). Jonathan Eisen, a microbiologist at the University of California, Davis, regularly announces awards for unnecessary additions to the scientific vocabulary on his blog (recent winner: CircadiOmics, for genes involved in daily circadian rhythms).

Botanist Hans Winkler had no idea what he was starting back in 1920, when he proposed the term 'genome' to refer to a set of chromosomes. Other 'omes existed even then, such as biome (collection of living things) and rhizome (system of roots), many of them based on the Greek suffix '-ome' — meaning, roughly, 'having the nature of'. But it was the glamorization of 'genome' by megabuck initiatives such as the Human Genome Project that really set the trend in motion, says Alexa McCray, a linguist and medical informatician at Harvard Medical School in Boston, Massachusetts. "By virtue of that suffix, you are saying that you are part of a brand new exciting science."

Researchers also recognize the marketing potential of an inspirational syllable, says Eisen. "People are saying that it's its own field and that it deserves its own funding agency," he says. But although some 'omes raise an eyebrow — museomics (sequencing projects on archived samples) and the tongue-in-cheek ciliomics (study of the wriggling hairlike projections on some cells) — scientists insist that at least some 'omes serve a good purpose. "Most of them will not make sense and some will make sense, so a balance should be in place," says Eugene Kolker, chief data officer at Seattle Children's Hospital in Washington, and founding editor of the journal *Omics*. "If we just laugh about different new terms, that's not good."

Ideally, branding an area as an 'ome helps to encourage big ideas, define research questions and inspire analytical approaches to tackle them (see 'Hot or not'). "I think -ome is a very important suffix. It's the clarion call of genomics," says Mark Gerstein, a computational biologist at Yale University in New Haven, Connecticut. "It's the idea of everything, it's the thing we find inspiring." Here, *Nature* takes a look at five up-and-coming 'omes that represent new vistas in science.

INCIDENTALOME

Several years before high-throughput sequencing made personal genomes a reality, Isaac Kohane, who studies medical informatics at Boston Children's Hospital, coined the term 'incidentalome' as a warning. The sheer quantity of available genetics information, he predicted in a 2006 article[1], would one day pose a challenge to medicine.

The term stems from 'incidentaloma', radiologists' slang for an asymptomatic tumour that shows up when doctors scan a patient for other complaints. The incidentalome describes the equivalent in human genome analyses: genetic information that no one was looking for. A search for the genetic cause of hearing loss in a child, for example, could turn up hints of future heart problems or a heightened risk of cancer. But who should be told what, and when? In an era in which more and more human genomes are being sequenced, the US National Human Genome Research Institute in Bethesda, Maryland, calls the question of what to tell individuals about their own DNA "one of the knottiest ethical issues facing genomics researchers".

A study last year[2] revealed the extent of the dilemma. It polled 16 genetic specialists about mutations implicated in 99 common genetic conditions that might show up in large-scale sequencing, whether or not a doctor was looking for them. For some 21 conditions or genes, including well-known sequence variants associated with certain cancers and a heart irregularity, all 16 specialists recommended informing adult patients. But only ten would do the same for Huntington's disease — an untreatable, fatal condition — and there was relatively little consensus on more obscure mutations, or what to tell parents when the variant showed up in a child's sequence.

The biggest problem with the incidentalome is that no one knows what most sequence variants — and there are more than 3 million in every human genome — mean for health. Wendy Chung, a clinical geneticist at Columbia University in New York, is developing ways to help research participants and patients to choose which genetic results they want to learn. She is also measuring the behavioural and psychosocial impacts of the information. "If you ask people what they want to know about their DNA sequences, everyone initially either says everything, or nothing," says Chung. "When people are thoughtful, there are shades of grey."

As clinical sequencing gains popularity, the definition and scale of the incidentalome is blurring. Geneticists should expect these hard-to-handle results, says Holly Tabor, a bioethicist at Seattle Children's Hospital. "It's somewhat misleading to say that there are incidental results from a genome study. You know that they will be there."

"BY VIRTUE OF THAT SUFFIX, YOU ARE SAYING THAT YOU ARE PART OF A BRAND NEW EXCITING SCIENCE."

PHENOME

Human genomes are now easy to come by. What's missing are phenomes: thorough, exact descriptions of a person's every physical and behavioural characteristic. Researchers most want to know about the portion of the human phenome related to disease: facial abnormalities, limb deformities, whether and how people were diagnosed with depression. And they want those descriptions in a form that computers can read — the better to see how such phenotypic traits might relate to genomes. "I do not know of another word or phrase with which we can say this better," says Peter Robinson, a computational biologist at the Charity University Hospital in Berlin, who is working to standardize such physical descriptions.

Phenome projects are already under way for mice, rats, yeast, zebrafish and the plant *Arabidopsis thaliana*. In the most systematic efforts, scientists knock out genes one by one, then carefully put organisms through a battery of measurements and physical tests to find out how genes shape physical form, metabolism and behaviour. Such comprehensive data cannot be had for human genes, but some clinical researchers hope to pull together a partial resource by carefully collecting patient data.

Even for 'Mendelian' diseases, known to be caused by a single mutated gene, matching up disease and gene is challenging. Of more than 6,000 rare, heritable disorders, fewer than half have been pinned to a genetic cause. One of the hardest parts is finding enough patients with such conditions, which may occur in fewer than one person in one million. "We could probably solve the majority of Mendelian disorders with an unknown cause if we had access to enough well-phenotyped cases," says Michael Bamshad, a geneticist at the University of Washington in Seattle.

But how to compile those cases? Many research and disease communities already have their own long-standing informatics tools and vocabularies to describe fine phenotypic details of various disorders. The challenge lies in getting these resources to work together. If one clinician enters 'stomach ache' and another 'gastroenteritis', patients with very similar symptoms may not get grouped together, explains Richard Cotton, a geneticist at the University of Melbourne in Australia.

In November last year, Cotton was among the scores of interested parties who came together in San Francisco, California, for a meeting called 'Getting ready for the Human Phenome Project'. The major aim of the meeting was to make the exchange of phenotypic data easier. A consortium that focuses on rare diseases, called Orphanet, is leading efforts to get clinicians and

NATURE.COM
To hear more debate on 'omics, listen to the podcast at:
go.nature.com/dd4ibo

scientists to agree on 1,000–2,000 standard terms — such as 'short stature', which may also be categorized as 'decreased body height', 'height less than 3rd percentile' and 'small stature'. "If you agree on the terms, no matter what form you have, we can all be talking about apples and apples and apples," says Ada Hamosh, a clinical geneticist at Johns Hopkins University School of Medicine in Baltimore, Maryland.

Other researchers are trying to unlock the often idiosyncratic information in electronic medical records so that computer algorithms can comb them and classify common phenotypes automatically. "The data are ugly and sparse, and the magic — the science — is turning that dross into gold," says Kohane.

> "I'M MORE EXCITED ABOUT THIS TECHNOLOGY THAN I'VE BEEN ABOUT ANYTHING IN A LONG, LONG TIME."

INTERACTOME

Biology's central dogma is essentially a parts list. DNA codes for RNA, which codes for protein. That may give you three basic 'omes (genome, transcriptome and proteome), but life happens only because these parts work together. A neuron fires and a cell divides or dies because molecules interact. The interactome describes all of those molecular interactions. And in terms of complexity, it is a king of the 'omes. Just considering one-on-one interactions for 20,000 or so proteins generates 200 million possibilities.

That scope is not daunting to researchers such as Marc Vidal. Before he retires, the 50-year-old systems biologist at the Dana-Farber Cancer Institute in Boston hopes to see a first, rough draft of all the interactions that the genome encodes. Actually, he would be happy with a subset, a catalogue of all the proteins that come together in pairs. "That's what we've been doing for the past 20 years, and we're almost there now," he says.

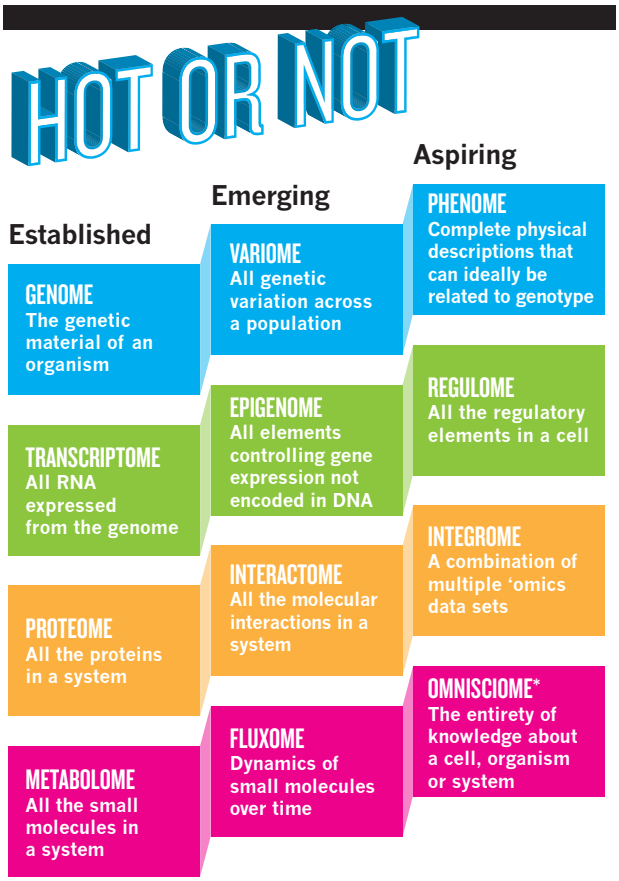
By 'almost there' Vidal means that his and a few other labs have observed 10–15% of human protein–protein interactions, based on studies of cells genetically engineered to generate a signal when a pair of proteins comes together. Other researchers have been pursuing the same goal by plucking proteins from crushed cells and seeing which others come along for the ride, scouring the literature and making computational predictions based on protein shapes and the behaviour of related molecules.

It has helped that, more than a decade after the first large-scale interactome study[3], researchers are finally starting to get a handle on which observed interactions are real and which are artefacts. Making that distinction requires hunting for the same interaction using multiple techniques[4]. But lists do not need to be complete to be useful — and biologists are already beginning to consult the interactome.

Haiyuan Yu, a systems biologist at Cornell University in Ithaca, New York, tested about 18 million potential protein pairs and combed established databases for interactions, eventually identifying 20,614 interactions between 7,401 human proteins. For around one-fifth of these interactions, the team also got a good sense of what parts of these proteins made contact[5]. Yu and his colleagues showed that disease-causing mutations are more likely to be at these points of contact than elsewhere in the proteins. For example, the blood disorder Wiskott–Aldrich syndrome is caused by mutations in a protein called WASP — but only by mutations located in an area that interacts with a second protein called VASP. Patterns that make no sense in terms of genes, says Yu, can become clear when considered in terms of interactions.

Vidal believes that increasingly sophisticated information can be layered into the interactome. First will come fleshed-out basic networks: lists of proteins and their binding partners, ideally annotated by cell types. Next will come descriptive data, such as how long interactions last, the conditions necessary for them and the parts of proteins that make contact.

Vidal imagines a day when clinicians diagnosing a patient will consider not only their genome, but the consequences of all their sequence variants on the interactome — not to mention the influences of the interactome on the phenome. Genomes, after all, are generally static, says Trey Ideker, a systems biologist at the University of California, San Diego. "The sequence is not perturbed by drugs, tissues or other conditions. Interactomes are."



**Nature*'s proposed addition to the scientific nomenclature.

TOXOME

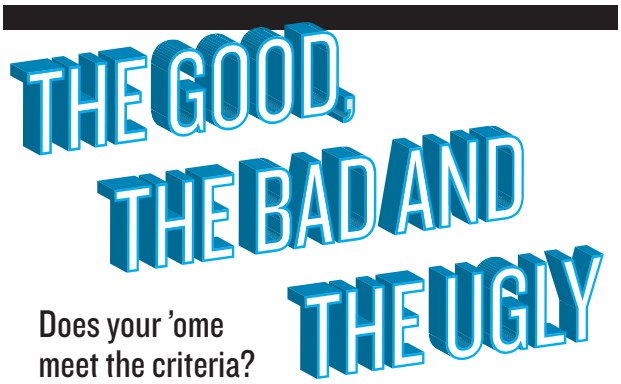
Thomas Hartung wants to learn all the ways a small molecule can hurt you. To do so, he has organized the Human Toxome Project, funded with US$6 million over five years from the US National Institutes of Health, plus extra support from the Environmental Protection Agency and the Food and Drug Administration. The -ome suffix, Hartung says, suited the scale of his goal: a description of the entire set of cellular processes responsible for toxicity. "The toxome is very similar to the Human Genome Project because it establishes a point of reference," says Hartung, a toxicologist at Johns Hopkins Bloomberg School of Public Health in Baltimore.

Toxicity testing in animal studies costs millions of dollars for every compound that enters human trials, yet animal tests sometimes fail to predict toxicity in humans. More than one in six drugs are pulled for safety problems that are discovered during human

trials. Hartung says that the toxome could help to lay out a series of straightforward cell-based assays that could replace animal tests — and perhaps improve on them. Knowing which toxicity-related processes a compound triggers could also help scientists to tweak promising new drugs or industrial molecules into less-harmful versions.

To start with, Hartung wants to expose cells to toxic chemicals and then monitor their metabolomes (the set of all small molecules in the cell) and their transcriptomes. He hopes to piece together the details of pathways in human cells that disrupt hormone signals, poison liver cells, break the heart's rhythm or otherwise endanger people's health. The total number of pathways, Hartung believes, will be perhaps a couple of hundred — a manageable amount for testing toxicities.

The project is still in its early days — making sure that the same assay yields the same results in different labs. Eventually, however, those pathways could be used in cell-based assays to serve as bellwethers of toxicity. "We'd know if we triggered one of those pathways that something bad would happen, and we'd know what that adverse event would be," says David Jacobson-Kram, who evaluates ways to predict toxicity at the Food and Drug Administration in Silver Spring, Maryland. He warns that a molecule that seemed harmless to cells in culture might behave differently in the body — for example if the liver converted it to a toxin. Nonetheless, he says, the toxome project could save time, money and animals. "Do I think this paradigm has promise?" he asks. "Absolutely."



THANKS TO JONATHAN EISEN, MICK WATSON AND ALEXA MCCRAY

INTEGROME

The key to unravelling biology's greatest mysteries depends less on inventing new 'omes, says Kolker, than on combining those that are already there. "One approach won't solve it," he says. Enter the integrome: information from all the 'omes thrown into one pot for an integrated analysis, along with any other relevant data for good measure. "That's the real deal, it's going to be more and more important," says Kolker.

Consider Google Maps. Separate lists of petrol stations, restaurants and street names are far less useful than one map showing that a particular petrol station is on the same street as a particular restaurant. But many conventional 'omics studies stop at list-making — genes, proteins or RNA transcripts. These can ignore networks and so may not reveal, for example, that changes in disparate genes actually converge on the same pathway.

Ideker has shown that it is possible to analyse disparate 'omics data automatically[6]. He created software that interrogated four collections of such data for patterns, and then used the results to work out independently what the relevant genes were doing. Not only did the software recapitulate parts of existing genome resources (for instance, identifying components of cellular machinery that help to dispose of spent proteins), but it started filling in gaps by finding similar patterns of organization for genes with unknown functions. "We trolled the transcriptome and interactome data and inferred the entire hierarchical structure of the components in a cell," says Ideker. "I'm more excited about this technology than I've been about anything in a long, long time." Such algorithms will not supplant human data curators, but they can pick up patterns that would be missed by humans or text-mining software that extracts relationships from published papers, he says. "Cells don't speak English; they speak data."

Last year, Michael Snyder, a geneticist at Stanford University in California, published his personal integrome[7] (although he called it an "integrative personal omics profile" — and others dubbed it the narcissome), combining data for his genome, transcriptome, proteome and metabolome (see *Nature* http://doi.org/hrq; 2012). The genomic profile revealed that Snyder had a risk variant for diabetes; during the study he was diagnosed with the disease and fought off two viral infections, which were reflected in increased activity of genes associated with inflammation. The 'omes also revealed changes in pathways not previously associated with diabetes or infection, says Snyder. "Had you only followed transcriptome or proteome, you would have only got part of the picture."

Gerstein agrees that integrated data sets are the way forward. "The future is going to be putting these things together in networks to understand personal genomes," he says. The word 'integrome', however, just doesn't sit right with him. "What is an integrome? The whole of all integrations? I don't think so." Integrate is a verb, he explains. "Most of the other 'omes are collections of nouns."

McCray has some rules of thumb for what constitutes a useful 'ome word: one that is meaningful, sounds pleasing and is easily understood by an educated audience (see 'The good, the bad and the ugly'). But it is unlikely that many scientists will take notice of the rules. The proliferation of words simply reflects the pace of the science, says McCray. Language typically changes slowly, but the rapid spread of the -ome and -omics suffixes is "recapitulating in a decade what normally takes a half century. It speaks to the intense interest and funding in the field." ■

Monya Baker *is a reporter for* Nature *in San Francisco, California.*

1. Kohane, I. S., Masys, D. R. & Altman, R. B. *J. Am. Med. Assoc.* **296,** 212–215 (2006).
2. Green, R. C. *et al. Genet. Med.* **14,** 405–410 (2012).
3. Uetz, P. *et al. Nature* **403,** 623–627 (2000).
4. Venkatesan, K. *et al. Nature Meth.* **6,** 83–90 (2009).
5. Wang, X. *et al. Nature Biotechnol.* **30,** 159–164 (2012).
6. Dutkowski, J. *et al. Nature Biotechnol.* **31,** 38–45 (2013).
7. Chen, R. *et al. Cell* **148,** 1293–1307 (2012).