BIOINFORMATICS

We the curators

Two groups describe wiki platforms for community-based curation of gene annotations or biological pathways.

The phenomenon that is Wikipedia, a free online reference resource that allows users to edit or create articles, has revolutionized the way many people obtain encyclopedic information. Wiki-based platforms can also be very useful for sharing scientific information. In one recent report, Andrew Su of the Genomics Institute of the Novartis Research Foundation and his colleagues describe their efforts to build up a comprehensive gene wiki within Wikipedia (Huss *et al.*, 2008). In another report, Bruce Conklin of the University of California, San Francisco, Chris Evelo of Maastricht University and their colleagues introduce WikiPathways, a community-based tool for biological pathway curation (Pico *et al.*, 2008).

Su and his colleagues saw great potential in using Wikipedia for community-curated gene annotations as a resource for both researchers and students. To rapidly build up a gene wiki, they 'seeded' Wikipedia with
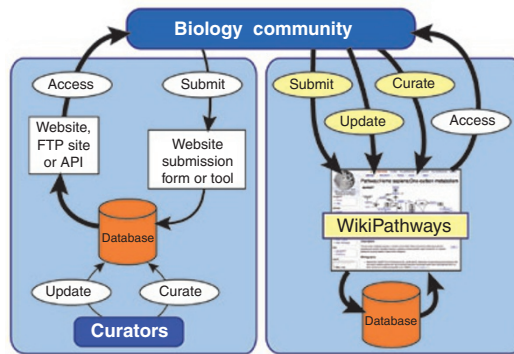

Figure 1 | Two models for managing biological data. The traditional database model (left) and the wiki-based model (right) are shown. FTP, file transfer protocol; API, application programming interface. Reprinted from *PLoS Biology* (Pico *et al.*, 2008).

7,500 new gene stub pages, based on data from Entrez Gene. "If you create minimally useful content you will attract readers; … the hope is that some percentage of those readers will actually stay to make an edit," explains Su. "It can be as trivial as fixing a typo, to summarizing a paper from the literature." This often generates a positive feedback loop; as pages gain more useful content they

attract more readers and more editors. Indeed, Su and his colleagues observed a substantial amount of editing to the stub pages even before they published a paper reporting their efforts.

Conklin and his colleagues thought that a wiki platform also had great potential for capturing biological pathway information, which is difficult to synthesize from static images in publications and from information scattered across different databases. Because such information is often visual rather than text-based, they amended the open-source software behind Wikipedia to make WikiPathways, which incorporates a new visual pathway editor. However, explains Conklin, "we are maintaining as much of the Wikipedia software as possible so that we can use their solutions for dealing with diverse communities as well as efficient curation methods."

Since WikiPathways has gone live, Conklin and his colleagues have already observed strong participation from diverse pathway communities. "There are communities that are already formed but didn't have a forum to

GENOMICS

RECOMBINATION: IT TAKES FOUR

Researchers use tetrad analysis and high-density oligonucleotide tiling arrays to generate a high-resolution map of meiotic recombination events in budding yeast.

During meiosis, genomic sequences are shuffled through homologous recombination, generating diversity in the genome. Combining classic yeast genetics and modern genomics, Lars Steinmetz's team at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany charted these meiotic recombination events at unprecedented resolution.

Steinmetz's group mated two budding yeast strains that differed genome-wide at more than 52,000 nucleotide positions and subsequently induced meiosis in them. They then scored meiotic recombination events in the resulting sets of 4 spores (tetrads) using the single-nucleotide polymorphisms as landmarks. To pinpoint the genomic positions of these events, Steinmetz's group turned to Affymetrix to generate high-density tiling arrays covering the genomes of both yeast strains. The high marker density and tight 4-base-pair spacing of the perfectly matched oligonucleotide probes on these arrays were especially important to accurately map non-crossovers, which tend to be extremely small. "With this technology, we would be able to invert the traditional relationship between markers and recombination events," explains Steinmetz, "[and perhaps] uncover something new about how recombination happens."

To identify individual meiotic crossovers and non-crossovers, Steinmetz's team turned to Wolfgang Huber's group at the EMBL European Bioinformatics Institute in Cambridge, UK, who developed a program, ssGenotyping, that automated these calls. Existing microarray-based genotyping tools just did not have sufficient accuracy to do the job.

"Non-crossovers are by their very nature short; typically they are seen only on one spore," says Huber, explaining how non-crossovers can easily be mistaken for microarray artifacts. "So, that was our challenge, to separate genotyping errors from the true [non-crossover] gene-conversion events." To meet these challenges, Huber's team implemented a semi-supervised clustering approach that takes into account the correlation of neighboring sequences (probes). They verified the accuracy of a subset of their genotyping calls by sequencing.

The maps generated had 20-fold greater resolution than the existing budding yeast map generated as a community effort using classic approaches and 360-fold greater resolution than the human map recently generated by Jonathan Pritchard and Molly Przeworski's teams (University of Chicago) using the Affymetrix 500K array. This is due to a greater density of markers and the tight spacing between oligonucleotide probes on the microarrays.

Tracing back the meiotic recombination events to their source, the initiating double-stranded breaks (DSBs), and determining the

push their pathway data out into the world," says Conklin. Alexander Pico, lead author on the report, also notes that "for our internal team and our collaboration with [colleagues from] the Netherlands, WikiPathways has made it so much easier to edit and curate content." Evelo describes another important benefit: "Pathways created on WikiPathways are not just representations of knowledge; they can also be used for analysis of new results."

Using a wiki-based platform for managing gene or pathway data is very different from the traditional model, in which researchers submit their data to databases that are curated by a relatively small number of experts (**Fig. 1**). Neither Su's nor Conklin's teams intend their wiki platforms to replace traditional databases, but they both emphasize that wikis can serve as highly complementary resources. "Even the people behind the gene portals recognize that their crew of 50 people can't keep on top of all the literature; … essentially you're empowering a large community to serve as amateur curators," explains Su. However, he stresses that with a wiki platform, "at any moment there may be a misleading statement, or it may be incorrect or incomplete—it's an eternal work in progress," and as such, wikis are not meant to serve as primary sources.

Perhaps one of the main reasons the wiki format works so well for communicating scientific data is because it is so simple and requires little effort to use. In using a wiki platform, Pico says, "an average biologist who's quite busy doesn't notice that at the same time that they're helping themselves they're also helping everyone else."

Allison Doerr

RESEARCH PAPERS
Huss, J.W. *et al.* A gene wiki for community annotation of gene function. *PLoS Biol.* **6**, e175 (2008).
Pico, A.R. *et al.* WikiPathways: pathway editing for the people. *PLoS Biol.* **6**, e184 (2008.)

frequency at which they occur, Steinmetz's group noted good agreement with the DSB map generated by Michael Lichten's team (US National Institutes of Health), providing additional confidence in their approach. Taking a closer look at their maps, the team identified 8 out of 9 previously known so-called recombination hotspots. Centromeres, mating-type loci and ribosomal DNA were, as expected, largely cold (recombination-free).

Notably, DSBs in certain regions of the genome were repaired through crossover events and others as non-crossovers. "There isn't a 50/50 choice in how a DSB is repaired," explains Steinmetz. "We think the chromosomal context influences this decision, … although this is an area that remains to be further investigated."

What is next? Huber's group is working toward a revised version of ssGenotyping that can be used as a general genotyping tool in which mutations can be accurately detected using one reference genome. "We feel based on the performance in this experiment, array-based genotyping based on this platform is reasonable for the detection of single-nucleotide polymorphisms, very short insertions and deletions," says Huber, "providing the basis for rapidly and inexpensively screening novel mutations." Huber and Steinmetz's teams hope to apply their tool to study the budding yeast as it evolves.

Michelle Pflumm

RESEARCH PAPERS
Mancera, E. *et al.* High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–485 (2008).

GENOMICS

Mapping the human methylome

Using different reductionist approaches, two groups now map the human methylome at nucleotide resolution. Using reduced representation bisulphate sequencing, Meissner *et al.* charted the methylomes of embryonic stem cells and primary neuronal cells, providing key insights into stem cell differentiation. With the aid of a new computational tool called Batman, Down *et al.* used DNA methylation immunoprecipitation coupled to high-throughput sequencing to chart the methylome of human sperm.

Down, T.A. *et al. Nat. Biotechnol.* **26**, 779–785 (2008).
Meissner, A. *et al. Nature* **454**, 766–770 (2008).

PROTEOMICS

Mice on a SILAC diet

Stable isotope labeling with amino acids in cell culture (SILAC) is a metabolic labeling technique used to quantitatively compare proteomic differences in two cell populations via mass spectrometry. Krüger *et al.* now extend the SILAC method to a mammalian model organism: the mouse. They found that mice could be SILAC-treated for several generations without any detrimental effects, and the labels were incorporated into all organ proteomes.

Krüger, M. *et al. Cell* **134**, 353–364 (2008).

GENE TRANSFER

A new strategy for zinc-finger nuclease design

Zinc-finger endonucleases can be custom-engineered to induce gene targeting at a user-specified sequence in the genome, but the strategies available to engineer these potentially very useful tools still have limitations. Maeder *et al.* now present a strategy termed oligomerized pool engineering (OPEN) based on an archive of selected zinc-finger pools and demonstrate its utility by designing nucleases targeting 11 sites in five different genes.

Maeder, M.L. *et al. Mol. Cell* **31**, 294–301 (2008).

MOLECULAR LIBRARIES

A library of histone mutants

As a resource for chromatin researchers, Nakanishi *et al.* present a complete alanine-scanning mutant library of the four core histones in *Saccharomyces cerevisiae*. They used this library to discover 18 core histone residues that are essential for viability and normal growth as well as to determine regulatory residues required for histone H3 Lys4 methylation.

Nakanishi, S. *et al. Nat. Struct. Mol. Biol.* **15**, 881–888 (2008).

PROTEIN BIOCHEMISTRY

Proximity biotinylation

Fernández-Suárez *et al.* report a new concept for detecting protein-protein interactions in live cells using an enzyme-substrate pair, exemplified by attaching the enzyme biotin ligase (BirA) to one protein and an acceptor peptide, which is site-specifically biotinylated by BirA, to the interacting protein partner. If the proteins interact, the acceptor peptide will become biotinylated, which can be detected by streptavidin staining.

Fernández-Suárez, M. *et al. J. Am. Chem. Soc.* **130**, 9251–9253 (2008).