WEB WATCH

Uniting human variation

- http://projects.tcag.ca/variation

It's becoming increasingly clear that polymorphism in the human genome goes way beyond SNPs. More and more studies are now identifying large-scale genomic variants — which include inversions, deletions and copy-number variants — as important components of normal human genetic variation. The Database of Genomic Variants aims to put all the information from these studies in one place.

The database is continually updated with information from both experimental data produced in-house by the research groups that curate it and from published studies. Just pick a chromosome and you are presented with a list of known variants, organized by location. Alternatively, you can enter the name of your favourite gene or region of the genome and the database will tell you if any identified variants are associated with it.

Useful graphical representations allow you to visualize where each variant lies in relation to cytological bands, coding regions and segmental duplications. The database also tells you the frequency of each variant, the method that was used to identify it and the ethnic backgrounds of the individuals in whom it was located. Links to gene databases and original papers make further investigations straightforward.

Once you've identified a variant of interest, the database also allows you to access a genome browser that provides more information about the surrounding genomic region. Here you can look for a range of features, including CpG islands, gene deserts, SNPs and segmental duplications.

As it expands, the database should help to piece together the contribution of large-scale polymorphisms to the genetic individualities that make each of us different.

Louisa Flintoft

TECHNOLOGY

It's cheaper in the Picolab

Although recent advances in DNA-sequencing technologies have accelerated the analysis of genomes from various organisms, including humans, sequencing whole genomes is still a time-consuming and expensive process. Margulies and colleagues have now developed a quick and easy sequencing method by combining an emulsion system for DNA amplification and pyrophosphate-based sequencing (pyrosequencing) in picolitre-sized wells.

Large-scale sequencing projects are laborious — the cloning of DNA fragments into bacterial vectors and the amplification and purification of individual templates is followed by Sanger sequencing using fluorescence chain–terminating nucleotide analogues and either slab-gel or capillary electrophoresis. Driven by the desire to reduce time and cost, Margulies *et al.* have devised a scalable, highly parallel two-step sequencing approach. The first step involves shearing the genome and generating random libraries of 80–120 bp DNA fragments. Adapters are ligated to the fragments. These are bound to beads and captured in the droplets of an oil-emulsion mixture of a PCR reaction. PCR amplification in each droplet results in each bead carrying ten-million copies of a unique DNA template. In the second step, a modified pyrosequencing protocol is carried out, in which nucleotide incorporation is detected by the release of inorganic pyrophosphate and the generation of photons.

The reactions take place on a picolitre scale: the slides used contain approximately 1.6 million wells, with 480 wells mm^{-2} and a calculated well size of 75 pl. The authors chose the bacterial genomes of *Mycoplasma genitalium* and *Streptococcus pneumoniae* to test the throughput, accuracy and robustness of their approach and achieved consensus accuracies of roughly 99.9%.

As the authors indicate, this study points to a "mini-approach" for future high-throughtput functional studies, which shows that miniaturization is becoming the core theme of genomics research. What does the future hold for the *de novo* assembly of genomes that are more complex than bacteria? The development of new sequencing methods is probably required for complex mammalian genomes, and the authors have already started to work towards this goal.

Ekat Kritikou

References and links

ORIGINAL RESEARCH PAPER Margulies, M. *et al.* Genome sequencing in microfabricated reactors. *Nature* 31 July 2005 (doi:10.1038/nature03959)

FURTHER READING Sauer, S. *et al.* Miniaturization in functional genomics and proteomics. *Nature Rev. Genet.* **6**, 465–476 (2005)

WEB SITE

454 Life Sciences: http://www.454.com/index2.html

GENE NETWORKS

Testing the network

The recent expansion of gene-expression and protein-interaction data sets means that we can predict increasingly complex models of gene-regulatory networks. But how does one recognize the correct model? Trey Ideker and colleagues have devised an automated method to minimize the effort of testing and refining gene-network predictions.

Combining gene-expression, promoter-binding and protein-interaction data, the authors constructed models of numerous gene-regulatory pathways in yeast. In some cases interactions between components were ambiguous — that is, more than one model was possible. One way to validate a particular prediction is to make genetic deletions of network components and determine whether the resulting changes in gene expression are consistent with the model. The authors developed an automated method that prioritized the deletion experiments that would provide the most information about their ambiguous models.

Three of the most informative experiments related to the same model — that for the regulatory pathway downstream of the yeast *SW14* and *SOK2* genes. Analysis of the deletions that were indicated by the automated system confirmed two predicted regulatory pathways within this model, whereas a third was rejected. Furthermore, when the results of these deletion experiments were combined with data from a previous analysis of 273 single-gene knockouts that was used to create the original model, many previously ambiguous interactions were resolved.

How does this method compare with other experimental validation strategies? Alternative methods might be to prioritize the deletion of hubs — genes that participate in many interactions — or to delete genes at random. The authors' automated method is more effective than either of these approaches at removing ambiguity. The method can also estimate the number of future experiments needed to fully refine a model.

The method described here has its limitations — it would be more useful if it could deal with experiments that involve the deletion of more than one gene. Nonetheless, this approach provides a useful tool for validating networks and could be made more powerful by extending its scope to multiple deletions in the future.

Louisa Flintoft

References and links

ORIGINAL RESEARCH PAPER Yeang, C.-H. *et al.* Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol.* **6**, R62 (2005)

FURTHER READING Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet*. **5**, 101–113 (2004)

WEB SITE

Pathway validation tool: http://www.cellcircuits.org/Yeang2005