

## NEUROSCIENCE

## An enhanced view of the brain

**A digital atlas of enhancers active in the developing mammalian brain is available for exploration.**

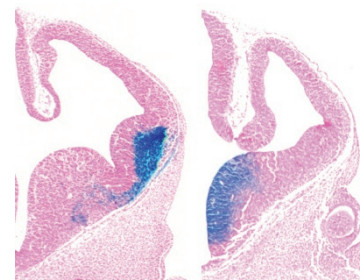
During development, each of the tissues in an organism exhibits spatially precise and perfectly timed gene expression patterns, like musicians in an orchestra playing a well-rehearsed symphony. But in the case of directing gene expression, the conductors work behind the scenes.

Gene expression in the cell is triggered by the binding of transcription factors to a gene's regulatory sequences.

Some of these regulatory sequences, also called enhancers, can be located hundreds of thousands of base pairs away from the promoters of their target genes, and often multiple enhancers control the expression of a given gene at once. Despite the critical role of enhancers in processes such as development, there is relatively little known about where exactly in the genome these sequences are located and when and where they are active in the organism.

Through the collaborative work of a team of functional genomics experts, developmental neuroscientists and computational biologists, the first high-resolution atlas of enhancer activity in the developing mammalian forebrain (a region comprising the cortex and the basal ganglia) has been released. The atlas includes the spatial patterns of activation of hundreds of enhancers during a critical time of brain development.

To generate this resource, the researchers, lead by Axel Visel and members of his laboratory at the Genomics Division of the Lawrence Berkeley National Laboratory, used several different



Enhancers exhibiting different activity patterns in the developing brain. Image courtesy of A. Visel.

## BIOINFORMATICS

## TUMORS HAVE THEIR DIFFERENCES

**Developing tools to detect the rare mutations found in tumors demands a lot of data and rigorous benchmarking.**

When a cell turns cancerous and begins to grow unchecked, genetic changes are to blame. Complicating the effort to find these rare changes is the fact that a single tumor is actually a collection of subclones, each with a unique spectrum of mutations that determines how aggressively the cell proliferates or responds to therapies.

“Calling mutations in cancer is more complicated than calling germline mutation,” says Gad Getz. Tumors are contaminated with normal cells, and parts of the cancer genome can exist as multiple copies, “so the signal for the mutation is diluted by all these normal cells and all these copies of the gene that are not mutated,” he says.

Getz directs the Cancer Genome Analysis Group at the Broad Institute that is responsible for analyzing cancer sequencing data generated at the Broad. He and his team developed the computational tool MuTect, a contraction of ‘mutation detection’ for finding somatic mutations in cancer including subclonal mutations at very low allelic fractions.

Calling mutations is not just a matter of tallying differences from the reference genome, because an overwhelming number of mutations result from sequencing and alignment errors and other technical artifacts. To develop a tool that reins in false calls while boosting sensitivity, the MuTect developers needed an innovative way to perform extensive benchmarking. They came up with the virtual tumor approach.

Benchmarking starts with a ground truth—a set of mutations that are known to be real by independent validation. But validated data are limited, and simulations do not capture all features of real data. As an alternative, the researchers created two sequencing libraries from a single tissue sample, labeling one ‘normal’ and the other ‘tumor’. Any mutations detected in the virtual tumor had to be false positives because they came from the same tube of DNA as the normal library.

techniques. They first performed chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) using an antibody for the enhancer-associated protein p300 from embryonic day 11.5 mouse brain tissue. p300 binds active tissue-specific enhancers. From the list of candidates, the researchers selected noncoding DNA sequences that were located far from transcription initiation sites, and this gave them 4,425 enhancer candidates.

Because not all the enhancers active in a tissue may be bound to p300 at a given time, they also performed computational analysis of genomic sequences based on high conservation between mouse and human and vicinity to genes known to be involved in forebrain development and function, which added 231 more sequences to the list of candidates.

The researchers then selected 329 candidate sequences, amplified them from human DNA, cloned them into an enhancer reporter vector and generated transgenic reporter mice. Of the 329 sequences, 105 exhibited reproducible activation patterns, and they analyzed these sequences in detail. The result was a vast collection of images showing spatially confined and beautifully defined enhancer activation patterns.

“When one looks at the diversity of patterns in the images it is pretty remarkable,” says Visel, “in many cases they don’t really match any single transcription factor expression pattern that we know.” The tens of thousands of high-resolution images in the collection can be explored through a dedicated website (<http://enhancer.lbl.gov/>).

There are several immediate applications of this resource. For one, it provides molecular tools of practical value to many researchers interested in targeting gene expression to subregions of the developing brain. The web-based accessible data also allow analysis of the biological properties of enhancers in more detail. Analysis of the sequences contained in these enhancers and their relation to the gene expression patterns they control will teach us the rules behind orchestrated genetic patterns that drive a tissue through development.

**Erika Pastrana**

#### RESEARCH PAPERS

Visel, A. *et al.* A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013).

To measure sensitivity, the researchers created a more sophisticated virtual tumor, by mixing DNA of relatives with a large number of known inherited differences—where, for example, a mother’s sequence differed from her daughter’s by a single base. The number of copies of maternal variants relative to those in the daughter, or allelic fraction, could be easily controlled to allow extensive testing. The virtual tumor “gives us a lot of data that we could actually do statistics on, and have small error bars on the specificity and sensitivity,” says Getz.

The variant calling engine of MuTect is derived from Bayesian statistics and classifies mutations as real or false by comparing the likelihood that they fit a reference model, which accommodates random sequencing errors or a variant model that takes the allelic fraction into account by estimating it from the sequencing data. Candidate mutations then face a gauntlet of filters, which remove false calls that arise from technical errors.

MuTect can detect mutations based on as few as three or four sequencing reads given a standard amount of sequencing, according to Getz. “So this is very, very sensitive. You can’t become much more sensitive than that,” he says.

What is the sensitivity good for? MuTect has been used to find new mutations in cancers with high confidence. In combination with Absolute, a tool from the Getz group that estimates tumor subclonal composition, MuTect allows researchers to look inside a single tumor mass and follow the genetic differences between cancer cells or reconstruct their evolution. The result is improved diagnostics and the potential to tailor treatments to the most aggressive subclones.

To develop a successful tool, Getz believes that there is no substitute to looking at the data. His group has enjoyed access to unusually extensive tumor sequence data, but he is happy to note that very large data sets are now being published by cancer genomics consortia.

**Tal Nawy**

#### RESEARCH PAPERS

Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).