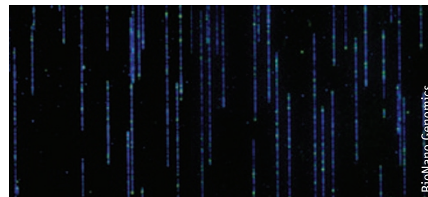
LAB-ON-A-CHIP

DNA: stretch for the camera

Using a special nanochannel chip, genome maps can be constructed from single DNA molecules in any laboratory.



Labeled DNA molecules line up like trains in a yard for genome mapping.

Optical DNA mapping may be the closest thing to taking out a ruler and measuring between the markers. DNA is typically pulled across a glass slide, then it is cut or hybridized to fluorescent probes at known sequences so that distances between spots or gaps can be calculated under the microscope. Although the method can be very effective, the difficulty of stretching the long and flexible DNA polymer uniformly has stymied its mainstream use.

Pui-Yan Kwok at the University of California at San Francisco, Ming Xiao at BioNano Genomics and their colleagues now show the potential for genome mapping to become widely accessible using a mass-produced chip that stretches DNA reliably for high-throughput automated measurements. The key to the chip is a set of 12,000 nanochannels that are each only 45 nanometers wide. "When DNA is in such a confined space, it cannot wiggle, so it's really stiff and straight, and the length is quite reproducibly set," says Kwok. "It's stretched to a degree that you can measure [distances between sequences] much better."

Maps derived from long single DNA molecules provide information that is difficult to obtain from sequencing and array-based genotyping. The authors genotyped molecules averaging >100 kilobases, which is substantially longer than reads generated by current sequencing methods. This capability is an advantage for repetitive DNA—regions which short sequencing reads are difficult to assign to unambiguously and for which physical maps provide effective scaffolds for genome assembly. The length also allows large structural variation to be reliably located. Hybridizing DNA to microarrays can find copy number changes

PROTEOMICS

MAKING PTMs A PRIORITY

Researchers describe an approach to zero in on post-translational modifications likely to have important regulatory functions.

Proteomics technologies are coming of age, allowing thousands of post-translational modifications (PTMs) to be identified in a single mass spectrometry run. But despite the recent impressive technology development, the really hard work—making sense of the data in the context of biology—is just beginning. "There's so much information one can generate on post-translational modifications nowadays since the 'mass specs' are so sensitive," says Nevan Krogan of the University of California, San Francisco. "A real challenge now is to take these long lists and identify [the PTMs] that we think are most likely to be functionally relevant, mutate the corresponding sites, and then functionally and phenotypically characterize those mutants."

With the help of a team of collaborators, Krogan's postdoc Pedro Beltrao devised a powerful computational framework to prioritize PTMs in terms of their functional relevance. The team compiled protein sequence data for more than 200,000 different PTM sites, including phosphorylations, acetylations and ubiquitinations, from 11 different species. They designed three classifiers to spot PTMs with potentially important functions: PTMs that are likely to regulate protein-protein interactions, PTMs that are likely to regulate domain activity and PTMs that are likely to cross-regulate each other. "We can't make a trained predictor," explains Beltrao. "We can only take biological intuition and [ask] what are the classes of PTMs that are likely to play a functional role based on what we know about the biology."

The researchers found that PTM sites predicted by the classifiers to be functionally important were indeed more highly conserved across the 11 species. They also identified conserved PTM sites within domain families that could be flagged as 'hot spots' of

© 2012 Nature America, Inc. All rights reserved.

npg

due to insertions and deletions, but unlike the physical maps, it cannot always tell where in the genome multiple copies originate, and it misses changes that do not affect copy number.

A challenge for nanochannel devices is to thread matted coils of DNA through a narrow opening. To accomplish this, an electric field first sends DNA past progressively smaller pillars on the chip, causing it to unwind like a ball of yarn in a pinball machine before entering the nanochannels. The chip can be flushed and replenished several times, allowing 3 gigabases of DNA to be processed in one experiment. To visualize genotypes, the authors labeled DNA prior to loading with a sequence-specific nicking enzyme followed by a polymerase that adds fluorescent nucleotides. After automated imaging, they applied clustering software to find overlaps between molecules with identically ordered labels, merging signals to give distances with a standard deviation of ~443 base pairs at 20× coverage. At this level of redundancy, essentially all missed or inappropriately labeled sites in single molecules are corrected.

As a proof of principle, the authors generated a map of the highly repetitive 4.7-megabase human major histocompatibility complex and used it as a scaffold to assemble the region from short-read sequence data. They detected structural and sequence differences between four parental DNA strands from two individuals, highlighting the fact that the method gives haplotype information. A haplotype is the unique string of sequences that originate from one parental chromosome. It is critical for many genetic analyses, but it cannot be derived directly from most sequencing and genotyping approaches.

Ultimately, it will be important to show that the method can tackle longer sequences. The groups have had success with human chromosome–length DNA and are working on high-density maps using multiple nicking enzymes and multiple fluorescent colors. "The mapping density [will] approach the sequencing level—not a single base but a few bases," says Xiao. Resolution can also be improved to around 100 base pairs using super-resolution imaging. There will be a lot to see as DNA lines up for the camera.

Tal Nawy

RESEARCH PAPERS
Lam, E.T. *et al*. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776 (2012).

© 2012 Nature America, Inc. All rights reserved.
npg

regulation. They followed up with mutational studies on a few predicted examples. For one, they showed that a phosphosite predicted to have a functional role in mediating a binding interaction was indeed crucial for the interaction of Skp1 with the Met30 F-box protein in the E3 ubiquitin ligase complex. They also discovered two regulatory hot spots in the heat shock protein 70 kDa (HSP70) chaperone domain family; when they mutated these phosphosites in HSP70 family member Ssa1, the mutant displayed growth defects under heat-shock conditions. "The rate-limiting step in all this is the more detailed targeted characterization of these PTMs; I don't think there's any way around that," says Krogan.

The work also demonstrates that many of the PTMs detected in large-scale mass spectrometry experiments are likely to be nonfunctional. Of course, as Krogan points out, some of the detected PTMs are probably artifacts from cell lysis, for example, when kinases go wild and start phosphorylating things they do not usually phosphorylate. However, other PTM sites may exist just because of their proximity to a functional site, or they may be artifacts of evolutionary dynamics, akin to pseudogenes in the genome. This underscores the need for approaches such as this to prioritize PTMs for experimental follow-up.

The team has created a database, http://ptmfunc.com/, which lists all the PTM sites predicted to be functionally relevant in this study. "People can look up their favorite protein and see if it's been reported to be modified, and this information, all of it is there," explains Krogan. Beltrao notes that this sort of analysis is expandable and that they are planning to include other types of classifiers in the future, such as PTMs that regulate localization signals. They are also working to expand the database to include additional species and other types of PTMs.

Allison Doerr

RESEARCH PAPERS
Beltrao, P. *et al.* Systematic functional prioritization of protein posttranslational modifications. *Cell* **150**, 413–425 (2012).