PSYCHOLOGY

Bias at the ballot box

"Could a seemingly innocuous factor, the type of polling location where people happen to be assigned to vote, actually influence how voters cast their ballots?" Jonah Berger and colleagues asked themselves this question, and went on to answer it with two types of study (J. Berger *et al. Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0711988105; 2008).

The first was an analysis of results from a general election held in Arizona in 2000, the ballot for which included a proposition to raise state sales tax from 5.0% to 5.6%, to increase education spending. Polling stations included churches, schools, community centres and government buildings.

Berger *et al.* predicted that voting in a school would produce more support for the proposition than voting in other places. Indeed it did, but not by much compared with other documented effects on voter choice such as order on the ballot paper. Nonetheless, the effect persisted through tests for various other confounding factors (for example, the possibility of a consistently different level of voter turnout at school polling locations).

The second study was a carefully run online experiment that also involved a proposed tax increase to fund schools. The 'voting environment' was manipulated by exposing participants to typical images of schools or control images. The upshot was the same, with the school images prompting greater (and apparently unconscious) support for the initiative than, for example, an image of an office.

All in all, the authors conclude that what they call contextual priming of polling location affects how people vote. They reasonably wonder whether such factors could, for example, influence voting in a church on such matters as gay marriage and stem-cell research.

But here's a thought. In the event

K. LAMARQUE/REUTERS/CORBIS

of science spending being on the political agenda, why not offer the lab as a polling station? But maybe dim that fluorescent lighting, and persuade all those bearded fellows in white coats to take the day off — or not, as the case may be.

Tim Lincoln

MOLECULAR BIOLOGY

Power sequencing

Brenton R. Graveley

Advances in DNA-sequencing technology provide unprecedented insight into the entire collection of four genomes' transcribed sequences. They herald a new era in the study of gene regulation and genome function.

Genomes are the blueprints of life: they contain all the information necessary to build and operate their hosts. But we still have much to learn about the language of DNA to interpret the billions of Gs, As, Ts and Cs, the DNA bases that spell out life. The information-containing portions of genomes are transcribed into two RNA classes: messenger RNAs, which are translated into proteins; and non-coding RNAs, which have regulatory and mechanical roles. So studying the transcribed portion of the genome — the transcriptome — significantly aids gene identification, as well as providing insight into the inner workings of the genome and the biology of an organism. Five recent papers[1–5], including one on page 1239 of this issue by Wilhelm *et al.*[1], describe how advances in DNA-sequencing technology can be harnessed to explore transcriptomes in remarkable detail.

The concept of sequencing large numbers of randomly selected mRNAs is not new. It forms the basis of the controversial, yet revolutionary, expressed sequence tag (EST) method[6], which was originally used to identify genes in the reference copy of the human genome. In this technique, genes are quickly identified through sequencing small fragments of large numbers of mRNAs. Although EST sequencing remains useful, it is relatively slow, requires considerable resources and generally cannot identify mRNAs that are expressed at low levels.

DNA microarrays are also powerful tools for transcriptome analysis. Particularly informative are tiling arrays, which are dotted with DNA sequences derived from defined intervals (for example, every 35 base pairs) throughout the genome. Fluorescently labelled RNA is then allowed to bind to the arrays, and the transcribed portions of the genome are identified by determining which DNA sequences pair with the RNA. But tiling arrays also have several shortcomings. First, they can be used only for organisms with known genome sequences. Second, their limited sensitivity, specificity and dynamic range (the ratio of the smallest to the largest fluorescent signal) make it difficult to identify low-abundance mRNAs and to distinguish between highly similar mRNA sequences. Finally, the number of DNA probes that fit on a microarray is limited, putting constraints on the minimum feasible genomic distance between the probes, and thus on the resolution at which a genome can be analysed.

Enter the trio of next-generation sequencing technologies — systems called 454 (from 454 Life Sciences), Solexa (from Illumina) and SOLiD (from ABI) — which can generate gigabases of sequence in a single experiment[7]. They differ from traditional sequencing methods in two ways. First, rather than sequencing individual DNA clones, hundreds of thousands (the 454 system) to tens of millions (Solexa and SOLiD) of DNA molecules are sequenced in parallel. Second, the sequences obtained are much shorter (25–50 nucleotides for the Illumina and ABI technologies, and 200–400 nucleotides for the 454 system) than those generated by traditional sequencing (typically more than 800 nucleotides). Matching these shorter sequences unambiguously to the reference genome is more difficult, but this is a relatively minor trade-off compared with the massive amount of total sequence generated using these technologies. The three sequencing systems have already revolutionized the study of chromatin structure, DNA-binding proteins, DNA methylation, genome organization and small RNAs[7]. But how useful they would be for studying transcriptomes was not known.

Five teams have now used a method called mRNA-Seq (Fig. 1, overleaf) to sequence, at various levels of detail, the transcriptomes of four organisms — the fission yeast *Saccharomyces pombe*[1], the budding yeast *Saccharomyces cerevisiae*[2], the plant *Arabidopsis thaliana*[3] and the laboratory mouse[4,5]. For the sequencing step, all except one of these groups used the Solexa system[1–4], and one team[5] used the SOLiD system. In each study, between 30 and 125 million sequences — 25–39 base pairs in length — were obtained. The most inclusive of these was performed by Wilhelm *et al.*[1], who generated 122 million 39-base-pair sequences for *S. pombe*, corresponding to nearly five gigabases of sequence or 250 equivalents of this organism's genome.

But how comprehensively do these analyses cover the known genes? In the one billion bases of sequence obtained for *S. cerevisiae*, only about 91% of the known genes are detected. By contrast, sequencing five billion bases of the *S. pombe* transcriptome, Wilhelm *et al.* identify 99.3% of known genes. So although 'moderate' sequencing of the transcriptome can quickly detect most genes, identification of all genes