

## GENOMICS

## Regeneration writ large

The assembly of genome sequences for the flatworm *Schmidtea mediterranea* and the salamander *Ambystoma mexicanum* will provide insights into the remarkable regenerative characteristics of these two organisms. SEE ARTICLES P.50 & P.56

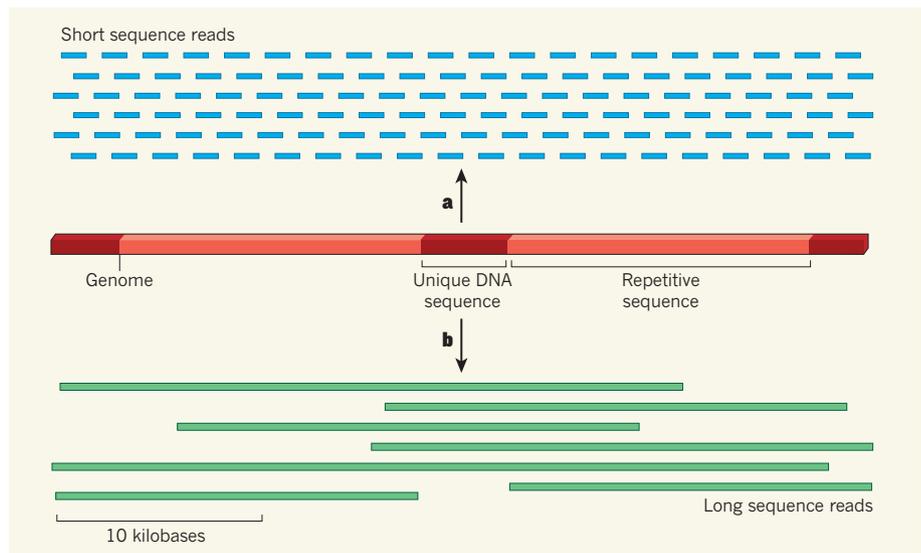
G. PARKER FLOWERS & CRAIG M. CREWS

Although humans have a limited ability to regenerate after injury, other animals can perform extraordinary regenerative feats. Small flatworms called planarians can regrow their entire bodies, even after being minced into hundreds of pieces. Many species of salamander can regrow whole limbs. For centuries, researchers have sought to understand these abilities, the instructions for which are encoded in DNA. In two papers in *Nature*, Grohme *et al.*<sup>1</sup> and Nowoshilow *et al.*<sup>2</sup> respectively report the genomes of a planarian (*Schmidtea mediterranea*) and a salamander, the axolotl (*Ambystoma mexicanum*). These studies mark a crucial step towards understanding regeneration.

The genome of *S. mediterranea* is composed of about 800 million bases spread over 4 chromosomes<sup>3</sup>, which makes it much smaller than the human genome. Even so, until now, the quality of the genome assemblies available for this species has been poor compared with that for other model organisms with larger genomes, such as mice, chickens and zebra fish, owing to problems with large-scale assembly of the planarian genome.

Imagine reconstructing a genome comprising hundreds of millions or even billions of bases from a jumbled library of individual sequence reads, each 500 bases or fewer in length. This is the challenge that has faced researchers using both traditional and next-generation sequencing methods. In theory, if there are enough overlapping individual sequences, a computer algorithm can stitch them together to recreate the genome. But the *S. mediterranea* genome contains abundant repetitive sequences, including virus-like sequences, such as retrotransposons, that have integrated repeatedly and replicated within the genome over the planarian's evolutionary history. These sequences are difficult to distinguish from one another using short samples (Fig. 1). Consequently, previous genome assemblies<sup>3,4</sup> based on short-read sequencing consisted of more than 100,000 different DNA fragments.

By contrast, Grohme *et al.* (page 56) used a long-read, single-molecule real-time (SMRT) sequencing platform to sequence the *S. mediterranea* genome. Using this strategy,



**Figure 1 | Assembling repetitive genomes.** The genomes of the planarian *Schmidtea mediterranea* and the axolotl *Ambystoma mexicanum* contain many, often long, tracts of repetitive sequences, separated by shorter unique sequences, including those containing protein-coding regions. **a**, Previously, these genomes were sequenced using short-read sequencing methods, which produce many DNA sequences that often cover only repetitive regions, making it difficult to identify their relative positions in the genome. **b**, Two groups<sup>1,2</sup> have used a method that allows for longer sequence reads (more than 10 kilobases), many of which include both repetitive and unique sequences. This enabled them to assemble genomes for *S. mediterranea* and *A. mexicanum*.

they obtained reads that averaged about 15,000 bases in length — longer than most of the individual fragments from earlier assemblies<sup>3,4</sup>. They stitched together the sequences by using a computer algorithm called MARVEL, which Grohme *et al.* and Nowoshilow *et al.* developed specifically to improve assembly of long reads from repetitive genomes. This approach bridges many of the assembly gaps caused by repetitive sequences, and produced a genome assembly with fragments more than 1 million bases long on average, which they then ordered into larger scaffolds of around 4 million bases. After this proof of principle, the approach could be applied to a much larger complex genome — that of the axolotl.

Nowoshilow *et al.* (page 50) used SMRT sequencing and MARVEL to assemble the 32-billion-base axolotl genome, which is ten times the size of the human genome. Almost two-thirds of the axolotl genome is made up of repetitive elements, many of which are more than 10,000 bases long. The authors assembled

sequence reads into fragments with a median length of 218,000 bases, which they compiled into scaffolds averaging about 3 million bases. The length and quality of these scaffold sequences are impressive, and demonstrate that some of the most complex genomes can be assembled using this approach.

The group estimates that axolotls have about 23,000 protein-coding genes — slightly more than humans but fewer than found in Grohme and colleagues' planarian assembly. The non-coding sequences within and between the genes are vastly larger than in humans and other vertebrates, mainly because of the expansion of repetitive elements. The group also identified microRNA sequences and genes that are missing in reptiles, birds and mammals, and whose expression was highly enriched in cells of the regenerating limb. Whether any of these candidate genes are crucial players in regeneration will be an interesting subject for future research.

Next, Nowoshilow and colleagues provided

evidence that *Pax3* — a gene essential for development in many animals — is absent in axolotls. They speculate that a related gene, *Pax7*, might compensate for this absence. Indeed, when the authors used gene-editing techniques to inactivate *Pax7*, mutant animals exhibited developmental abnormalities and muscle loss similar to those seen in mice lacking both *Pax3* and *Pax7*.

In a paper recently published in *Nature Communications*, Elewa *et al.*<sup>5</sup> found that this compensatory action of *Pax7* is probably restricted to a subset of regenerative salamanders. Their genomic study shows that both *Pax3* and *Pax7* are retained in another salamander, the Spanish ribbed newt (*Pleurodeles waltl*), which also has a portfolio of impressive regenerative abilities. Mutational analysis reveals that *Pax7* is not required for normal muscle development and regeneration in this newt, whereas *Pax3* is essential. The ability to investigate and compare the roles of genes in development and regeneration across species in this way marks the beginning of a new era of research in these models of regeneration.

Grohme and colleagues compared the genomes of *S. mediterranea* and other planarian species, and were unable to detect 124 genes that are essential in humans and mice. These genes include some involved in DNA repair and some that have essential roles in protecting against errors in chromosome segregation during cell division. How such organisms manage to thrive without components that have been regarded as essential for life in vertebrates is a fascinating question.

The new genome assemblies, when combined with the sudden ease of genetic manipulation using new genome-editing tools, will make it possible to do experiments that were previously unimaginable in model organisms such as planarians and salamanders. For example, consider the repetitive stretches of DNA that hampered the assembly of the current genomes. In both species, the main contributors to these repeats are retrotransposons. Previous studies<sup>6,7</sup> have suggested that retrotransposons contribute to important biological processes that shape embryonic development and stem-cell behaviour. Elewa *et al.*<sup>5</sup> found that retrotransposons are expressed in regenerating limbs in the Spanish ribbed newt, and might in turn regulate gene expression. Whether and how these elements have been co-opted to guide regeneration in various species are among the many exciting avenues of research that can now be explored. ■

**G. Parker Flowers** and **Craig M. Crews** are in the Department of Molecular, Cell and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA. C.M.C. is also in the Departments of Chemistry and Pharmacology, Yale University. e-mails: grant.flowers@yale.edu; craig.crews@yale.edu

1. Grohme, M. A. *et al.* *Nature* **554**, 56–61 (2018).
2. Nowoshilow, S. *et al.* *Nature* **554**, 50–55 (2018).
3. Robb, S. M. C., Gotting, K., Ross, E. & Sánchez Alvarado, A. *Genesis* **53**, 535–546 (2015).
4. [https://www.ncbi.nlm.nih.gov/assembly/GCA\\_000691995.1#/st](https://www.ncbi.nlm.nih.gov/assembly/GCA_000691995.1#/st)

5. Elewa, A. *et al.* *Nature Commun.* **8**, 2286 (2017).
6. Fort, A. *et al.* *Nature Genet.* **46**, 558–566 (2014).
7. Magiorkinis, G., Katzourakis, A. & Lagiou, P. *Trends Microbiol.* **25**, 876–877 (2017).

This article was published online on 24 January 2018.

## BREAST CANCER

# A rude awakening from tumour cells

**In women who have had breast cancer, drug treatments are often stopped five years after removal of the primary tumour. A meta-analysis shows that these individuals are still at risk of relapse.**

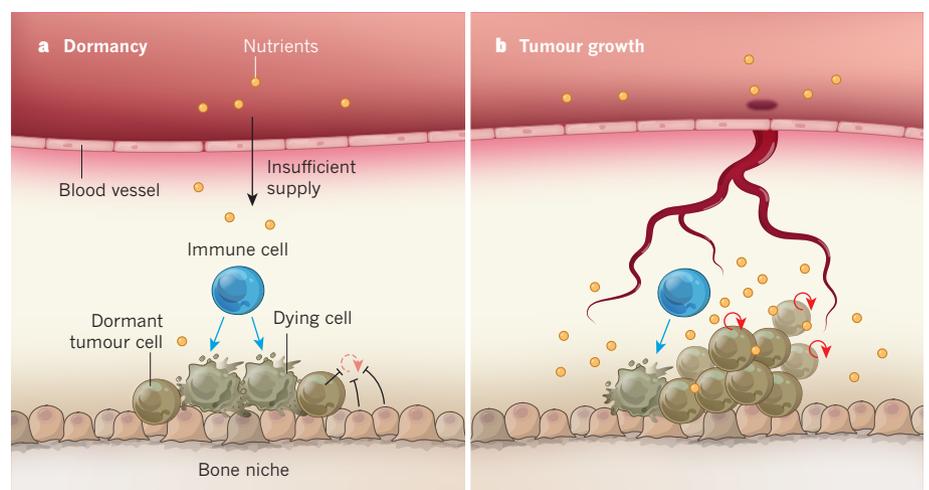
GIUSEPPE CURIGLIANO & FATIMA CARDOSO

The main aim of adjuvant therapy, which is given after an apparently successful primary cancer treatment, is to reduce the risk of local and distant metastatic disease relapse owing to residual breast-tumour cells that can persist for years or decades in a dormant state. Our knowledge of the biology of dormant residual disease is cripplingly limited. Writing in the *New England Journal of Medicine*, Pan *et al.*<sup>1</sup> examine rates of metastatic cancer spread in 62,923 women treated for breast cancer and given adjuvant therapy. The authors' findings provide a window on dormancy in this disease.

Pan and colleagues performed a meta-analysis of 88 trials involving women who

had ER-positive breast cancers — subtypes of breast cancer characterized by expression of the oestrogen receptor (ER). The women were all disease-free after five years of scheduled adjuvant endocrine therapy, which involves taking drugs that lower the activity of the ER. The beneficial effects of these treatments for preventing metastasis during the 5 years following diagnosis are not in doubt — instead, the authors' analysis was designed to determine the risk of late metastasis occurring between years 5 and 20, if adjuvant therapy is stopped after 5 years. They found that metastasis occurred at a steady rate for the 15 years after the end of the treatment period.

Remarkably, the most powerful determinants of the risk of recurrence were those originally used to grade the aggressiveness of the primary



**Figure 1 | Multiple mechanisms underlying tumour dormancy.** Pan *et al.*<sup>1</sup> report that women who take drugs for five years after their primary breast cancer has been removed are still at risk of a late relapse, because of tumour cells that have migrated to a secondary site such as the bone and adopted a dormant state. **a**, Tumour-cell dormancy can involve many factors, including: poor blood-vessel supply, which means that there is insufficient oxygen and nutrients for proliferation; ongoing surveillance by immune cells that kill some dormant tumour cells; and signals, both tumour-cell-intrinsic and from cells in the surrounding bone niche, that inhibit proliferation (dashed curly arrow). **b**, Relapse occurs when the balance between proliferation and these factors is lost, because of vessel growth, evasion of immune-cell activity or changes that prevent proliferation-blocking signals, leading to tumour growth.