medium to high confidence. Importantly, the studies found that using only 30 to 100 landmark genes was sufficient for accurately assigning the cells to the spatial areas.

Satija *et al.*[2] studied the late-stage blastula zebrafish embryo and defined a rather coarse-grained grid in the reference map that accommodated 40–120 cells per grid area. Indeed, this large-sized grid worked adequately for localizing cells in a zebrafish embryo as the tissues have low complexity. The algorithm (called Seurat) can also compensate for the stochastic noise in gene expression of single-cell RNA-seq profiles[1] on the basis of the correlation with co-expressed genes. Achim *et al.*[3] defined much smaller grid areas in the reference map corresponding to 11–174 grid areas per cell (i.e., subcellular resolution) and mapped cells with good precision using high-resolution fluorescent ISH atlases of the *P. dumerilii* larval brain. In general, both methods should be easy to implement in any laboratory. The approaches could also be extended by replacing the qualitative ('on' and 'off') scoring of ISH signals in the reference maps with a quantitative scoring to account for graded differences in gene expression.

Resolving spatial gene expression is an important step toward understanding the complexity of heterogeneous tissues. Single-cell analyses alone can identify and study the cellular heterogeneity, but in many cases the biological meaning of the heterogeneity will become apparent only when information about single cells is combined with their spatial information. It is possible to see where subpopulations are located within tissue structures, which could provide clues as to their potential cellular functions. Also, spatial information allows gene expression of interacting cells to become visible, enabling a greater understanding of signaling networks in a specific region.

Numerous complementary efforts have been devised to obtain spatial gene expression resolution[4]. Developments in ISH technology (sequential fluorescent ISH, or FISH) enable more genes to be analyzed in one experiment and at higher sensitivity (single-molecule FISH)[4]. RNA sequencing of specific *in vivo* cells (transcriptome *in vivo* analyses, or TIVA[5]) can provide high-quality cellular profiles, but is difficult to scale up to the complexity of all cell types within a tissue. Alternative methods such as laser-capture microdissection or tomo-seq[6] can be used for spatial transcriptome profiling, but not with single-cell resolution. Recently, progress has been reported in sequencing of RNA molecules *in situ*[7,8]; however, this technology is currently limited to the sequencing of a few RNA bases from a limited number of genes.

As they make use of the existing wealth of ISH databases, the methods described here are likely to be a relatively cost effective means to add a spatial dimension to single-cell RNA-seq data. Nevertheless, both approaches are limited to tissues for which corresponding ISH atlases are already available, and naturally the precision of the spatial mapping is limited to the resolution of the ISH data lodged in the repositories. In cases where only smaller ISH atlases exist, it is always possible that new ISH experiments can be designed and landmark genes identified from analyses of the single-cell profiles. Although it seems fairly easy to map back cells of distinct cell types (for which many marker genes exist) to specific locations, it could prove more difficult to correctly place cells that are highly similar in expression or scattered in the structure. Furthermore, it goes without saying that this technology will be applicable only to well-ordered tissues, as there are no consistent ISH images for more variably structured tissues, such as solid tumors.

Finally, it will be interesting to use these methods[2,3] to map single cells onto more complex three-dimensional (3D) tissues. Reference maps in 3D could be generated from existing resources, such as the Allen Brain Atlas[9] or be generated *de novo* using RNA-seq on cryosectioned tissue slices (e.g., tomo-seq[6]). It may even be possible to base 3D reference maps on RNA or protein image analyses in unsectioned transparent whole tissues using approaches like CLARITY[10]. In any case, these studies highlight that combining imaging and single-cell sequencing technologies can now be applied to decode gene expression networks in complex tissue structures.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Sandberg, R. *Nat. Methods* **11**, 22–24 (2014).
2. Satija, R., Farrel, J.A., Gennert, D., Shier, A.F. & Regev, A. *Nat. Biotechnol.* **33**, 495–502 (2015).
3. Achim, K. *et al. Nat. Biotechnol.* **33**, 503–509 (2015).
4. Crosetto, N., Bienko, M. & van Oudenaarden, A. *Nat. Rev. Genet.* **16**, 57–66 (2015).
5. Lovatt, D. *et al. Nat. Methods* **11**, 190–196 (2014).
6. Junker, J.P. *et al. Cell* **159**, 662–675 (2014).
7. Lee, J.H. *et al. Science* **343**, 1360–1363 (2014).
8. Ke, R. *et al. Nat. Methods* **10**, 857–860 (2013).
9. Lein, E.S. *et al. Nature* **445**, 168–176 (2007).
10. Chung, K. *et al. Nature* **497**, 332–337 (2013).

Unraveling the fabric of polyploidy

Andrew H Paterson & Jonathan F Wendel

Draft genome sequences of tetraploid Upland cotton and other polyploids promise insights into genome evolution and cotton fiber biology.

Although uncommon in animals, ancient (paleo-) polyploidy is ubiquitous in plants, with recent (neo-) polyploidy being common in cultivated crop species. In this issue, Zhang *et al.*[1] and Li *et al.*[2] report draft genome sequences of allotetraploid ($2n = 4x = 52$) *Gossypium hirsutum* (Upland cotton). Analysis of these genome sequences sheds new light on why cotton polyploids have higher yields and improved fiber productivity compared with their diploid progenitors[3] (**Fig. 1**). More generally, sequencing of polyploid genomes provides new information about crops that have been among the most challenging to study.

Whole genome sequencing projects must produce a sufficient number of sequence 'reads' covering each nucleotide in the genome to distinguish occasional errors from the true sequence by parsimony. Using high-fidelity but costly Sanger sequencing, 6- to 8-fold redundancy is often sufficient to differentiate DNA sequences that are up to ~99.8% similar. Cheaper, massively parallel sequencing methods compensate for higher error rates by making it affordable to obtain the 100× or higher coverage of most nucleotides in a genome, which is needed to resolve similar sequences.

Polyploidy, which is the inhabiting of a nucleus by multiple genomes, adds an additional layer of complexity during genome assembly. All flowering plants are paleopolyploid[4], with at least one (and often several) genome duplication(s) in their evolutionary history. In many plants, including wheat, potato, strawberry, tobacco, canola and cotton, neopolyploidy (recent merger of two or more similar genomes) has also occurred. A further complication in genome assembly of neopolyploid species is that the two newly co-resident genomes may interact, exchanging and even 'overwriting' genetic material either reciprocally or nonreciprocally, thus obliterating portions of their independent evolutionary

Andrew H. Paterson is at the University of Georgia, Athens, Georgia, USA, and Jonathan F. Wendel is at Iowa State University, Ames, Iowa, USA.
e-mail: paterson@uga.edu

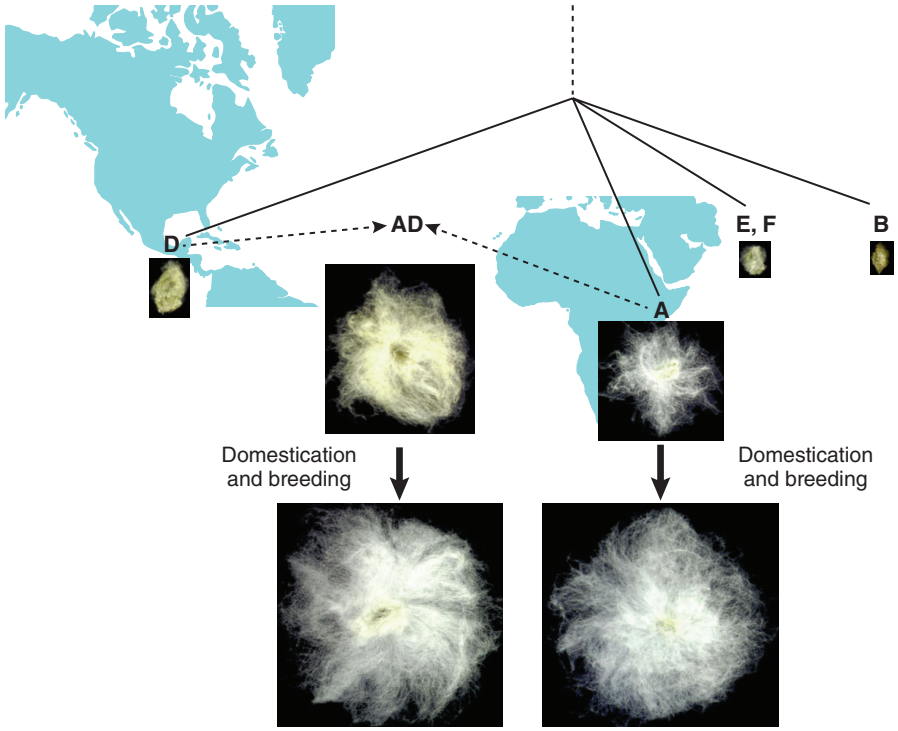

Figure 1 Evolution of tetraploid cotton. Dashed line in evolutionary tree indicates extinct ancestor. Spinnable cotton 'lint' fibers evolved in an ancestor of Old World 'A genome' cottons, following their divergence from the B, E and F genomes about 5–10 million years ago. All known tetraploid cottons can be traced to a single common ancestor that formed about 1–2 million years ago in the New World, through transoceanic dispersal of an 'A genome' ancestor resembling *Gossypium herbaceum* (diploid) followed by mating with an indigenous 'D genome' species resembling *Gossypium raimondii* (diploid) (dashed line arrows). Domestication and breeding has resulted in tetraploid cottons (*G. hirsutum* and *G. barbadense*) that consistently have higher yield and quality of fiber than improved diploids. Outline maps from http://www.nationsonline.org/.

history and the very information that is needed to distinguish them from one another.

The comparative framework offered by phylogenetics provides a key to the assembly of polyploid genomes. For such polyploids as cotton (*Gossypium*), crucifers (*Brassica*) and wheat (*Triticum*), sequencing of extant diploid progenitors has provided reference genomes that lack the complications of neopolyploidy. Diploid progenitor genome sequences enable the tracing of the provenance of genome segments in a polyploid. For example, most assembled scaffolds of the canola (*Brassica napus*) genome were matched specifically to one but not both of its progenitors[3]. Furthermore, diploid reference genomes can be used as a template for analyzing massively parallel resequencing data from tetraploids, to reveal signatures of selection and homoeologous gene exchanges[5–7]. A shortcoming of resequencing methods is that they have only limited power to dissect features associated with longer DNA regions such as chromosome structural variations, which have classically been thought to play an important role in cotton speciation[8]. Therefore, full draft genome assemblies of tetraploid cotton are necessary to identify tetraploid-specific features potentially relevant to future genetic improvement.

Before rejoining 1–2 million years ago (MYA)[9], the A- and D-genome progenitors of *Gossypium* (cotton) polyploids experienced 5–10 million years of independent evolution on opposite sides of the Atlantic Ocean, providing sufficient divergence to permit unambiguous sequence assembly[6]. This divergence, and previously published reference genomes for both progenitor species (*Gossypium arboreum* and *Gossypium raimondii*), facilitated the assembly of two whole genome sequences of the same tetraploid cultivar (Texas Marker Stock (TM)-1)[1,2]. Researchers in both teams used paired bacterial artificial chromosome (BAC)-end sequences and genetic mapping to increase whole genome contiguity to a level approaching that of the draft diploid sequences[10,11], although these draft genomes are not as accurate as the *Gossypium raimondii* D-genome reference sequence[6]. Genes annotated in the two sequences differ by 12%, and curiously the gene-rich annotation (76,943) also contains a higher percentage (67.2%) of repetitive DNA[9].

What do these two new assemblies reveal about how tetraploid cottons differ from their diploid progenitors? Gene loss in tetraploid cotton, which was previously difficult to detect[12,13], is found in both assemblies at lower rates than in other polyploid plants but at higher rates than in the progenitor diploids, and is more frequent in the physically larger but gene-poorer A_t (228 of 32,032) compared with the D_t subgenome (141 of 34,402). Indeed, the physically larger A_t subgenome has generally been more extensively altered than the D_t subgenome[8], with its rich repetitive DNA making it more prone to repeat-mediated recombination than the smaller and less repetitive D genome. It remains to be investigated whether this gene loss is ancient and present in all tetraploid cotton species, or recent and perhaps related to yield or fiber biology of elite cultivars. One intriguing hypothesis is that there may also be an ecological basis for genome evolution, with polyploids sharing the New World habitat of the D-genome progenitor, whereas A-genome cottons are of Old World origin and presumably only existed in the New World briefly by means of long-distance dispersal[9]. Changes in gene expression were known to be rapid, dramatic and tissue-specific in both synthetic and natural polyploid cottons compared with diploid lines[14]. The comprehensive genome-wide RNA-seq analysis of Zhang *et al.*[1] shows that gene expression bias for duplicated genes is widespread, frequently in the 20–40% range, depending on the tissue.

These tetraploid assemblies also confirm and extend previous resequencing studies. Extensive homoeologous exchanges between A-derived tetraploid (A_t) and D-derived tetraploid (D_t) subgenomes were abundant in a progenitor of polyploid cottons[6], and are now known to be among the earliest events that differentiate other new polyploids from their progenitors[5]. However, exchanges of sizable chromosome blocks could also be inferred in one TM-1 assembly[2]. Likewise, signatures of selection in 'resequenced' tetraploids[6,7] were added to by the new assemblies[1,2] to further enrich lists of candidate genes for ongoing investigation of fiber evolution and improvement, as well as other adaptations of cotton to cultivation.

Both tetraploid assemblies support prior 'dating' of polyploid cotton formation to about 1–2 MYA[7], but do little to clarify an intriguing question about cotton paleo-evolution. All analyses agree that cotton shared a paleo-hexaploidy (3× multiplication) occurring in a core eudicot ancestor, but differ as to whether a subsequent polyploidization was tetraploidy only 13–20 MYA[9–11], or deca (5×)- or dodeca (6×)-ploidy closely following the cotton-cacao divergence ~60 MYA[4]. The '20 MYA tetraploid school'[11] offers dates for cotton-cacao divergence that are

only half the 60 million years of age of eumalvoideae pollen fossils[15] and overlook 5- to 6-fold duplication in better-assembled parts of their genome such as a segment of chromosome 13 corresponding with segments of chromosomes 1, 5, 8, 10 (2) and 12 (ref. 11, Supplementary Fig. 7), essentially nullifying their case. However, better clarity as to the exact course of cotton paleoevolution would be of much interest, and may result from sequencing of additional genomes that last shared common ancestry with *Gossypium* in the intervening time.

Many other allopolyploids, such as wheat, oats and peanut, in which two or more subgenomes retain sufficient divergence to permit unambiguous sequence assembly[6], may benefit from characterization similar to that carried out for canola[5] and cotton[1,2]. A more daunting challenge is genome assembly of 'autopolyploid' genomes such as sugarcane and alfalfa, that have homologous chromosome sets of 4, 6, 8 or more members that are regularly homogenized by continuous pairing and recombination. Most of the crops that sustain humanity, and indeed the planet's dominant terrestrial vegetation (angiosperms), are polyploid, meaning that understanding how the merger of two (or more) genomes leads to emergent 'transgressive' properties, such as improved yield, is of crucial importance. Projects that tackle assembly of polyploid plant genomes, such as those reported herein[1,2] and elsewhere recently[5], pave the way to a better understanding of the basis of the evolution and improvement of complex traits in crop species.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Zhang, T. *et al. Nat. Biotechnol.* **33**, 531–537 (2015).
2. Li, F. *et al. Nat. Biotechnol.* **33**, 524–530 (2015).
3. Jiang, C.X., Wright, R.J., El-Zik, K.M. & Paterson, A.H. *Proc. Natl. Acad. Sci. USA* **95**, 4419–4424 (1998).
4. Bowers, J.E., Chapman, B.A., Rong, J. & Paterson, A.H. *Nature* **422**, 433–438 (2003).
5. Chalhoub, B. *et al. Science* **345**, 950–953 (2014).
6. Paterson, A.H. *et al. Nature* **492**, 423 (2012).
7. Guo, H. *et al. Genetics* **197**, 1153–1163 (2014).
8. Stephens, S. *Genetics* **34**, 627–637 (1949).
9. Wendel, J.F. & Cronn, R.C. *Adv. Agron.* **78**, 139–186 (2003).
10. Li, F. *et al. Nat. Genet.* **46**, 567–572 (2014).
11. Wang, K. *et al. Nat. Genet.* **44**, 1098–1103 (2012).
12. Liu, B., Brubaker, C.L., Mergeai, G., Cronn, R.C. & Wendel, J.F. *Genome* **44**, 321–330 (2001).
13. Rong, J. *et al. Heredity* **108**, 500–506 (2012).
14. Yoo, M.J., Szadkowski, E. & Wendel, J.F. *Heredity* **110**, 171–180 (2013).
15. Carvalho, M.R., Herrera, F.A., Jaramillo, C.A., Wing, S.L. & Callejas, R. *Am. J. Bot.* **98**, 1337–1355 (2011).

Islet implantation in a pocket

Jonathan S Bromberg

Exploiting the foreign body reaction creates a prevascularized space that accommodates and supports transplanted islets.

A long-standing barrier to the successful transplantation of pancreatic islets to cure insulin-dependent type 1 diabetes is difficulty finding an ideal site of implantation. Human islet transplantation is done by infusing the cells into the portal vein of the liver, a procedure accompanied by technical, hematologic, inflammatory, vascular and trophic complexities that diminish short- and long-term success (**Fig. 1**). Proposed sites alternative to the portal vein have not been successful or scalable to large animals or humans. In this issue, Shapiro and colleagues[1] describe a simple approach to create a subcutaneous site for islet implantation in mice. They implanted clinically approved nylon catheters subcutaneously for a few weeks, inducing a vascularized pocket. Islets were infused into the pockets, where they survived, increased both their glucose responsiveness and their insulin output over time, and successfully treated diabetic recipients almost as well as traditional murine islet transplantation to the kidney subcapsular space. The technique overcomes technical and physiologic limitations of portal infusion, is amenable to large animals, could be adapted to clinically relevant drugs and platforms (e.g., growth and angiogenic factors, slow-release formulations, biocompatible scaffolds), and opens new opportunities for imaging, biopsy, local immunomodulation and stem cell treatments.

Although human islet transplantation has been attempted for many decades, the 'Edmonton protocol' has substantially reinvigorated the field, yielding short-term islet allograft survivals >90%; previous eras had achieved success rates of only <5% (ref. 2). Longer term follow-up has revealed limited durability of graft survival, with most experiences demonstrating <50% at 5 years[3]. Barriers to durability involve several major factors, including the quality and quantity of transplanted islets, immediate engraftment, and long-term survival and function[4].

The current report addresses the major areas of early implantation, engraftment and survival. The only approved and successful technique for clinical islet transplantation is portal vein infusion, which embolizes islets throughout the liver. Immediate complications of bleeding and thrombosis due to the technique are well-known but manageable. More difficult to control is the immediate loss of a large percentage of the inoculum due to the coagulation, inflammation and innate immunity that are stimulated by the infusion[4]. Delayed vascularization and early deficiency of oxygen and trophic factors further hinder survival[4]. Preventive measures have focused on inhibitors of inflammation, coagulation, immunity and cell death pathways but have had limited success. Supplementary growth and vascular factors have had limited efficacy in preclinical models[5].

Much effort has focused on developing encapsulation techniques and devices to shroud islets and prevent some adverse events[6]. Although successful in small animal models, applicability to large animals is poor due to inflammation, thrombosis and fibrosis caused by device materials and lack of scalability to support a sufficient islet mass for clinical feasibility. Another focus is development of alternative implantation sites (e.g., brain, thymus, stomach, pancreas, spleen, omentum, kidney, muscle, dermis, bone marrow and testis)[7] to obviate immunity and inflammation while permitting desirable vascular and trophic effects. This focus has met with only limited success because the same barriers of inflammation, immunity, scalability, vascularization and ease of administration persist.

Another limitation to the portal vein and many other sites is the inability to image, biopsy or retrieve the implanted islets. Indeed, islet transplantation is unique because imaging and biopsy—the current mainstays of the entire rest of the fields of organ, tissue and hematopoietic stem cell transplantation—are not used. These modalities are routinely and frequently required to diagnose rejection, infection, drug toxicity, recurrent disease and other complex events. Islet transplantation stands alone in being unable to adopt these important measures, which provide critical information for clinical management. Thus, an easily accessible site that concentrates the islets in a confined space would be a major advance in the field.

In the current report, Shapiro and co-workers take advantage of the foreign body reaction to all implantable devices, which induces varying amounts of inflammation, neovascularization

Jonathan S. Bromberg is in the Departments of Surgery and Microbiology and Immunology, Center for Inflammatory and Vascular Diseases, University of Maryland School of Medicine, Baltimore, Maryland, USA.
e-mail: jbromberg@smail.umaryland.edu