castration-resistant metastatic tumors, suggesting that dysregulation of the androgen receptor pathway could be an early event in prostate tumorigenesis. Genes for other transcriptional regulators, such as the transcriptional initiation component *MED12*, are also mutated in both stages of disease.

ETS fusion–negative tumors

Over half of prostate cancers bear translocations involving the genes of ETS family transcription factors, such as *ERG* and *ETV1* (refs. 8,9). This has led to the idea that ETS fusion–positive tumors form a distinct subgroup[6], but the molecular basis of disease in ETS fusion–negative cancers is less clear. Some ETS fusion–negative tumors have translocations in genes in the Raf kinase pathway[10] or have outlier expression of *SPINK1* (ref. 11), but the genetic mechanisms remain unexplained in many of these cancers. Barbieri *et al.* find that mutations in *SPOP*, present in 6–15% of primary tumors, are mutually exclusive from ETS translocations and likely explain a significant fraction of ETS fusion–negative cancers. *SPOP* encodes an E3 ubiquitin ligase component[12] and is also mutated in colorectal cancer in sequences encoding the distinct BTB dimerization domain. *SPOP* mutations in prostate cancer exclusively affect the substrate-binding MATH domain, implicating SPOP as a tumor suppressor, as the mutated protein binds more weakly to substrate *in vitro*. However, *SPOP* copy-number loss is rarely, if ever, observed in prostate cancer, raising the possibility that these mutations may confer *de novo* gain of function. Alteration in the *CHD1* gene encoding a chromatin-remodeling factor defines another ETS fusion–negative class of prostate cancer[2], as this gene was previously found to be mutated or rearranged in three out of seven prostate cancer whole genomes sequenced[3]. Collectively, tumors with mutated *SPOP* and *CHD1* account for a substantial fraction of ETS fusion–negative prostate cancers.

With a fairly complete list of prostate cancer genome alterations spanning the full spectrum of disease now in hand, it is time to ask the critical question of whether this information can guide a more precise approach to the diagnosis and treatment of primary disease. Earlier data comparing copy-number alterations in primary and metastatic disease found that late-stage tumors resemble a subset of highly altered primary tumors and that altered copy number in primary disease was associated with a greater risk of relapse[5]. Comparison of copy-number alterations observed in primary versus castration-resistant disease by Barbieri *et al.* and Grasso *et al.*, respectively, supports the observation of increased copy-number alteration in prostate cancer, albeit without clinical data to conclude association with outcome (**Fig. 1**). The whole-exome mutation data from these new reports provide additional variables that should be examined in future studies of prognosis, together with mRNA expression profiles recently reported to correlate with outcome[7,13]. Future work may allow patients with high-risk disease to be treated more aggressively, perhaps with novel agents targeted at relevant driver lesions, whereas patients with low-risk disease might be watched for signs of progression without treatment. The wider lens provided by these reports therefore offers the hope of tailoring treatment of prostate cancer on the basis of genomic risk and the presence of driver lesions.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Barbieri, C.E. *et al. Nat. Genet.* **44**, 685–689 (2012).
2. Grasso, C.S. *et al. Nature* published online, doi:10.1038/nature11125 (20 May 2012).
3. Berger, M.F. *et al. Nature* **470**, 214–220 (2011).
4. Robbins, C.M. *et al. Genome Res.* **21**, 47–55 (2011).
5. Taylor, B.S. *et al. Cancer Cell* **18**, 11–22 (2010).
6. Demichelis, F. *et al. Genes Chromosom. Cancer* **48**, 366–380 (2009).
7. Cuzick, J. *et al. Lancet Oncol.* **12**, 245–255 (2011).
8. Tomlins, S.A. *et al. Nature* **448**, 595–599 (2007).
9. Shen, M.M. & Abate-Shen, C. *Genes Dev.* **24**, 1967–2000 (2010).
10. Palanisamy, N. *et al. Nat. Med.* **16**, 793–798 (2010).
11. Tomlins, S.A. *et al. Cancer Cell* **13**, 519–528 (2008).
12. Furukawa, M., He, Y.J., Borchers, C. & Xiong, Y. *Nat. Cell Biol.* **5**, 1001–1007 (2003).
13. Cuzick, J. *et al. Br. J. Cancer* **106**, 1095–1099 (2012).

Exploring the variation within

Evan Z Macosko & Steven A McCarroll

We usually think of an individual's cells as sharing the same genome. Challenging this notion, two new studies show that somatic mosaicism is common and can be an early herald of cancer.

What do the trillions of cells comprising an individual human have in common? Aside from the commensal organisms that colonize the body, the prevailing answer to this question has been the human genome. In normal parlance, as well as in the scientific literature[1], we speak of a genome sequence as belonging to a given person and existing in all of that individual's cells. But how accurate is this notion? Studies by Laurie *et al.*[2] and Jacobs *et al.*[3] in this issue report that our DNA changes in subpopulations of cells over time and that these changes may predict the subsequent development of cancer.

Evan Z. Macosko and Steven A. McCarroll are in the Department of Genetics at Harvard Medical School, Boston, Massachusetts, USA, and in the Program in Medical and Population Genetics at the Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
e-mail: mccarroll@genetics.med.harvard.edu

Over 100 years ago, Wilhelm Roux and August Weismann independently asserted that the particles of heredity are differentially apportioned during embryonic divisions to give rise to genetically different cells that have unique roles in development, an idea they called mosaicism[4,5]. Later studies largely rejected this hypothesis, showing that somatic tissues arise from the differential use of genes that are shared across all cells. With notable exceptions of specialized processes such as the V(D)J recombination that occurs in certain immune cells[6], examples of mosaicism have largely been limited to congenital genetic illness[7] and recognized cancer[8,9].

More common than we think

Recently, however, several studies have hinted that somatic mosaicism is more widespread than previously thought. One study of human embryos fertilized *in vitro* found that 70% incurred segmental imbalances postmeiotically[10]. Another study detected mosaicism in 1.7% of blood and buccal genomic DNA samples[11]. And earlier this year, somatic variation in blood samples from a twin cohort identified copy-number differences in 3% of twin pairs[12]. These results suggest that detectable clonal mosaicism occurs in a measurable proportion of adults over 50 years of age.

Just how pervasive is clonal mosaicism in the general population? In at least one tissue, the blood, the collection of samples for genome-wide association studies (GWAS) enables surveys of genomic mosaicism in large population

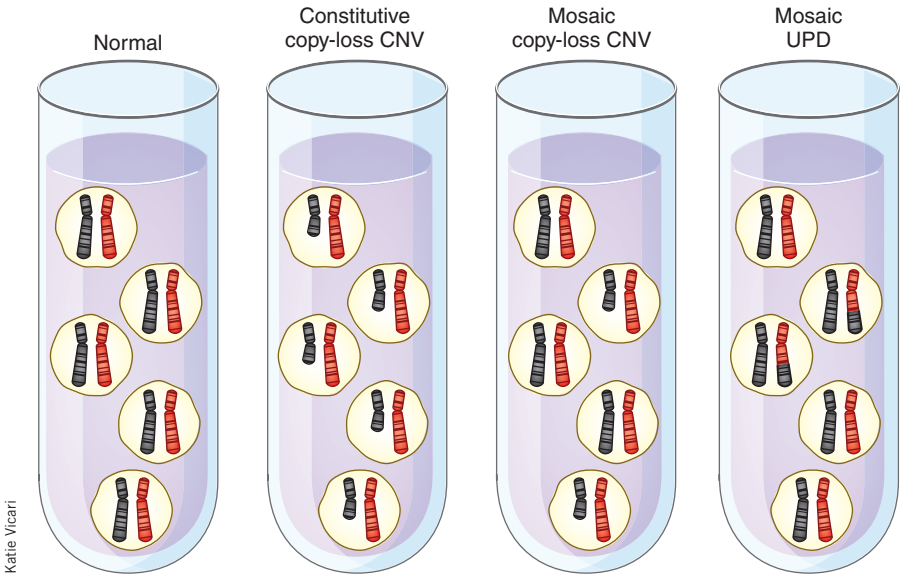

Figure 1 Populations of cells with normal karyotypes, constitutive chromosomal abnormalities or mosaic chromosomal anomalies. Paternal chromosomes are shown in black. Maternal chromosomes are shown in red.

cohorts. Moreover, the molecular technology used for GWAS (SNP arrays) is well suited to such analysis. Whereas constitutional copy-number variations (CNVs) involve integer changes in the representation of a chromosomal segment (jumps to 50% or 150% of the expected representation), mosaic gains and losses cause changes of smaller magnitude. These are made visible by statistical analysis of hybridization intensities and allelic ratios: long tracts of SNPs with altered hybridization intensities and allelic ratios other than 1:1 and 2:0 signal gains or losses of chromosomal segments in a fraction of the cells interrogated[2,3,13]. Segmental uniparental disomy (UPD), caused by mitotic recombination and other events that result in loss of heterozygosity without a change in copy number, can also be detected from allelic ratios. Using appropriate statistical methods, studies can now identify segmental losses, large segmental gains and UPD events that are present in as few as 5% of the cells interrogated (**Fig. 1**). The studies by Laurie *et al.* and Jacobs *et al.* found that detectable mosaicism is rare (<1%) in adults younger than 50 but that its prevalence increases to 2–3% in adults older than 70. This suggests that clonal populations of cells expanded to comprise a detectable fraction of blood cells in many individuals.

The studies by Laurie *et al.* and Jacobs *et al.* both found that these clonal expansions have an important medical consequence: individuals with detectable clonal mosaicism in blood had greatly increased risk of subsequently developing hematological cancer. These effects were large, with relative risk estimated at 10 by Laurie *et al.* and 35 by Jacobs *et al.*; the confidence intervals for these two estimates overlapped each other, and they excluded the null hypothesis (of no relationship) with overwhelming statistical significance. Some of the identified deletions and duplications affected genes previously implicated in leukemia and myeloma and are likely to have promoted clonal expansions within the blood.

Beyond identifying specific genes affected by these deletions and duplications, detectability of mosaicism in the blood may be an important marker for the disproportionate expansion of clonal populations of cells. Intriguingly, Jacobs *et al.* also report that certain solid tumors may occur at higher prevalence in individuals with detectable clonal mosaicism in their blood. This suggests a potential early developmental origin to some of these events or, alternatively, that genomic stability or the organism-wide management of clonal expansion may be compromised in some individuals.

These studies raise the possibility of many interesting research directions. One will involve exploring how these clonal expansions relate to developmental lineages. Enrichment of genetically aberrant clones and their characterization using approaches such as flow cytometry may help elucidate whether these cells are restricted to a specific myeloid or lymphoid lineage or exist across multiple germ layers.

On a clinical level, detectable mosaicism could represent a 'pre–pre-cancerous' state, a prelude to conditions such as myelodysplastic syndrome (MDS), considered by many to foreshadow leukemia[14]. In the way that MDS is now treated to prevent or forestall cancer, clinical investigations may come to indicate therapeutics (or careful monitoring, at least) for individuals harboring oligoclonality or detectable mosaicism. Screening tools based on the detection of oligoclonality and mosaicism may be another potential direction.

A new frontier

These studies begin to explore an exciting frontier in genetics research: the soma. The observation of substantial mosaicism in blood invites other questions. How widespread is mosaicism in other tissues? How substantial is the genome variation within an organism?

The study of mosaicism could, in principle, find itself on a scientific path much like that traversed recently by copy-number variation. Once viewed exclusively as a hallmark of cancer and congenital disease, deletions and duplications have come to be recognized as also part of the reservoir of human genome variation. Will we eventually see somatic mosaicism as part of the human condition? Current research strategies aimed at the detection of clonal expansions and large chromosomal events may tend to see the forms of mosaicism most associated with malignancy and disease. But an increasing ability to see smaller events and events in smaller populations of cells may eventually add new dimensions to this view. Recent innovations in single-cell technologies should enable a deeper understanding of genomic diversity in the soma.

Just as population genetics has helped us to understand the dynamics of natural variation in groups of individuals, our growing knowledge of somatic genetics may yield a new set of principles characterizing growth behavior—expansions, bottlenecks and selection—of the cells within an organism. We may even find that aspects of Roux and Weismann's early theory of mosaicism are less obsolete than we thought.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Wheeler, D.A. *et al. Nature* **452**, 872–876 (2008).
2. Laurie, C.C. *et al. Nat. Genet.* **44**, 642–650 (2012).
3. Jacobs, K.B. *et al. Nat. Genet.* **44**, 651–658 (2012).
4. Weismann, A. in *Essays Upon Heredity and Kindred Biological Problems* (eds. Poulton, E.B., Schonland, S. & Shipley, A.E.) (Oxford University Press, London, 1889).
5. Roux, W. *Z. Biol.* **21**, 411–524 (1885).
6. Hozumi, N. & Tonegawa, S. *Proc. Natl. Acad. Sci. USA* **73**, 3628–3632 (1976).
7. Youssoufian, H. & Pyeritz R.E. *Nat. Rev. Genet.* **3**, 748–758 (2002).

8. Rowley, J.D. *Cancer Invest.* **1**, 267–280 (1983).
9. Yunis, J.J. *Science* **221**, 227–236 (1983).
10. Vanneste, E. *et al. Nat. Med.* **15**, 577–583 (2009).
11. Rodríguez-Santiago, B. *et al. Am. J. Hum. Genet.* **87**, 129–138 (2010).
12. Forsberg, L.A. *et al. Am. J. Hum. Genet.* **90**, 217–228 (2012).
13. Conlin, L.K. *et al. Hum. Mol. Genet.* **19**, 1263–1275 (2010).
14. Tefferi, A. & Vardiman, J.W. *N. Engl. J. Med.* **361**, 1872–1885 (2009).

One gene's shattering effects

Kenneth M Olsen

A new study shows that three independent mutations in the *Sh1* gene, which encodes a YABBY transcription factor, gave rise to the non-shattering seed phenotype in domesticated sorghum. This same gene may have also had a role in the domestication of other cereals, including maize and rice.

The shift from freely shattering seeds, which easily fall off a plant at maturity, to non-shattering or reduced-shattering seeds represents a key transition during cereal crop domestication. Whereas wild grasses evolve under strong selection for the ability to disperse their seeds at maturity, domestication favors plants from which entire grain stalks can be efficiently harvested with minimal seed loss (**Fig. 1a**). However, once the shift to non-shattering grains occurs, the reproductive fate of a crop species becomes intimately tied to its human cultivators: subsequent crop generations depend on harvesting and resowing for their continued existence. In recent years, one of the major aims of crop domestication research has been to understand how non-shattering grains and other domestication traits evolved. Following pioneering work in the 1990s by John Doebley and colleagues in maize[1,2], studies in cereals and other crops have begun to resolve the genetic mechanisms underlying traits favored either during the initial stages of domestication (such as losses of seed shattering and dormancy) or during subsequent breeding for crop improvement (such as diversification in grain pigmentation and starch characteristics)[3].

Only recently, however, have enough species and traits been examined that we can begin to ask whether the same genes underlie the same traits in different crop species. In this issue, Jianming Yu and colleagues[4] make an important advance in addressing this question. Working in *Sorghum*, with comparative analyses in rice and maize, the authors identify the key gene responsible for non-shattering grains in domesticated sorghum, and they report evidence that orthologs of this gene may have contributed to the non-shattering phenotype in rice and maize as well.

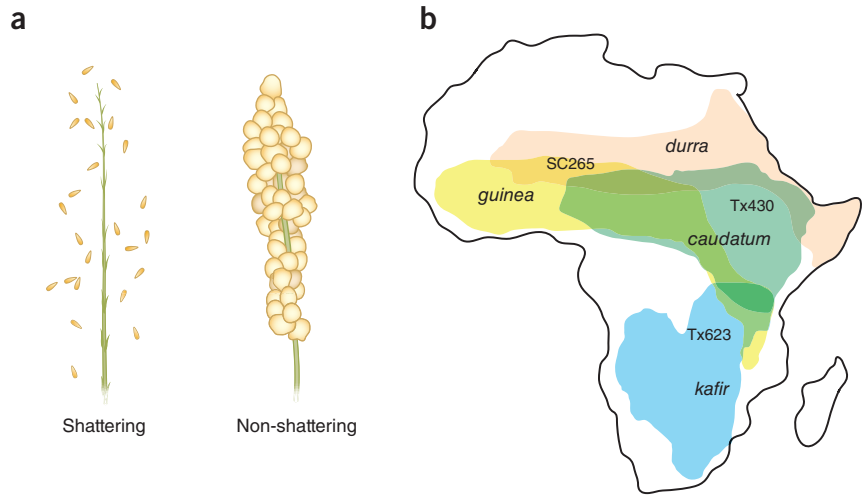

Figure 1 Multiple origins of non-shattering sorghum in Africa. (**a**) Shattering and non-shattering grains are characteristic of wild and domesticated varieties, respectively. After shaking, mature seeds scatter easily in shattering varieties. In contrast, seeds remain on the head of non-shattering varieties. (**b**) Approximate geographic distributions of sorghum varieties with *sh1* non-shattering alleles (SC265, Tx430, Tx623) among the major domesticated sorghum races in Africa: *durra* (light brown), *guinea* (yellow), *caudatum* (green) and *kafir* (blue). The *bicolor* sorghums (not shown) occur throughout the geographic range covered by the other crop forms.

Sorghum's multiple origins

Domesticated sorghum (*Sorghum bicolor*) provides a major source of calories for livestock and humans worldwide, with sorghum grain production ranking third among cereal crops in the United States and fifth globally. It is also considered to be an emerging bioenergy crop. Sorghum was domesticated in Africa, which still harbors a great diversity of cultivated forms. Five major morphological forms or 'races' have traditionally been recognized[5,6]: *caudatum*, originating from eastern Africa; *durra*, predominant in the Horn of Africa and other arid regions; *guinea*, most characteristic of western Africa; *kafir*, dominant in subequatorial eastern Africa; and the widely distributed *bicolor*, which includes both shattering and non-shattering varieties and is considered the most primitive form of the crop (**Fig. 1b**). A recent genome-wide analysis of SNP diversity confirmed the genetic distinctness of the four fully domesticated races[6], a pattern potentially consistent with multiple independent domestication events.

Whereas in most crops shattering is controlled by many interacting genes, shattering in sorghum is controlled by a single major-effect quantitative trait locus (QTL)[7,8]. In their new study, Lin *et al.*[4] use map-based cloning to identify the causal gene, *Sh1*, which they find encodes a YABBY transcription factor. By sampling a large and diverse sorghum collection, the authors provide evidence that the non-shattering phenotype can be accounted for by

Kenneth M. Olsen is at the Department of Biology at Washington University, Saint Louis, Missouri, USA.
e-mail: kolsen@wustl.edu