introduction of transgenic maize developed for pharmaceutical or other non-food purposes, and its impact on landraces[9,10].

Mexico needs to be able to define what kind of transgenic materials (for maize and any other relevant crop) it needs for its ecological, social and economic requirements. This responsibility must be carefully analyzed in order to provide farmers with adequate and necessary elements to help achieve a level of food security for the present and future of Mexican society, while conserving genetic diversity and helping develop adequately the social structures of the rural economy and society.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

DISCLAIMER
The manuscript reflects only the opinion of the authors and not the institution they represent.

Francisca Acevedo, Elleli Huerta, Caroline Burgeff, Patricia Koleff & José Sarukhán

Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México D.F., México.
e-mail: facevedo@conabio.gob.mx

1. Anderson, E. & Cutler, H.C. *Ann. Mo. Bot. Gard.* **29**, 69–86 (1942).
2. Kato, T.A., Mapes, L.M., Mera, L.M., Serratos, J.A. & Bye, R.A. *Origen y Diversificación del Maíz: una Revisión Analítica* (Universidad Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México, D.F. (2009).
3. Bellón, M.R. & Brush, S.B. *Econ. Bot.* **48**, 196–209 (1994).
4. Bellón, M.R. *et al. Diversidad y Conservación de Recursos Genéticos en Plantas Cultivadas, en Capital Natural de México, vol. II: Estado de la Conservación y Tendencias de Cambio.* (CONABIO, México, 2009).
5. Bourges, H. in *La Alimentación de los Mexicanos* (eds. Alarcón-Segovia, D. & Bourges, H.) 97–134 (El Colegio Nacional, México D.F., 2002).
6. Hernández-Xolocotzi, E. *Econ. Bot.* **39**, 416–430 (1985).
7. Pressoir, G. & Berthaud, J. *Heredity* **92**, 88–94 (2004).
8. Dalton, R. *Nature* **462**, 404 (2009).
9. Acevedo, F. *Nat. Biotechnol.* **22**, 803 (2004).
10. Acevedo G.F. *et al. La Bioseguridad en México y los Organismos Genéticamente Modificados: Como Enfrentar un Nuevo Desafío, en Capital Natural de México, vol. II: Estado de la Conservación y Tendencias de Cambio* (CONABIO, México, 2009).

Integrative genomics viewer

To the Editor:

Rapid improvements in sequencing and array-based platforms are resulting in a flood of diverse genome-wide data, including data from exome and whole-genome sequencing, epigenetic surveys, expression profiling of coding and noncoding RNAs, single nucleotide polymorphism (SNP) and copy number profiling, and functional assays. Analysis of these large, diverse data sets holds the promise of a more comprehensive understanding of the genome and its relation to human disease. Experienced and knowledgeable human review is an essential component of this process, complementing computational approaches. This calls for efficient and intuitive visualization tools able to scale to very large data sets and to flexibly integrate multiple data types, including clinical data. However, the sheer volume and scope of data pose a significant challenge to the development of such tools.

To address this challenge, we have developed the Integrative Genomics Viewer (IGV), a lightweight visualization tool that enables intuitive real-time exploration of diverse, large-scale genomic data sets on standard desktop computers. It supports flexible integration of a wide range of genomic data types including aligned sequence reads, mutations, copy number, RNA interference screens, gene expression, methylation and genomic annotations (**Supplementary Fig. 1**). The IGV makes use of efficient, multi-resolution file formats to enable real-time exploration of arbitrarily large data sets over all resolution scales, while consuming minimal resources on the client computer (**Supplementary Notes**). Navigation through a data set is similar to that of Google Maps, allowing the user to zoom and pan seamlessly across the genome at any level of detail from whole genome to base pair (**Supplementary Fig. 2**). Data sets can be loaded from local or remote sources, including cloud-based resources, enabling investigators to view their own genomic data sets alongside publicly available data from, for example, The Cancer Genome Atlas[1], 1000 Genomes[2] (http://www.1000genomes.org/) and ENCODE[3] (http://www.genome.gov/10005107) projects. In addition, IGV allows collaborators to load and share data locally or remotely over the internet.

IGV supports concurrent visualization of diverse data types across hundreds,

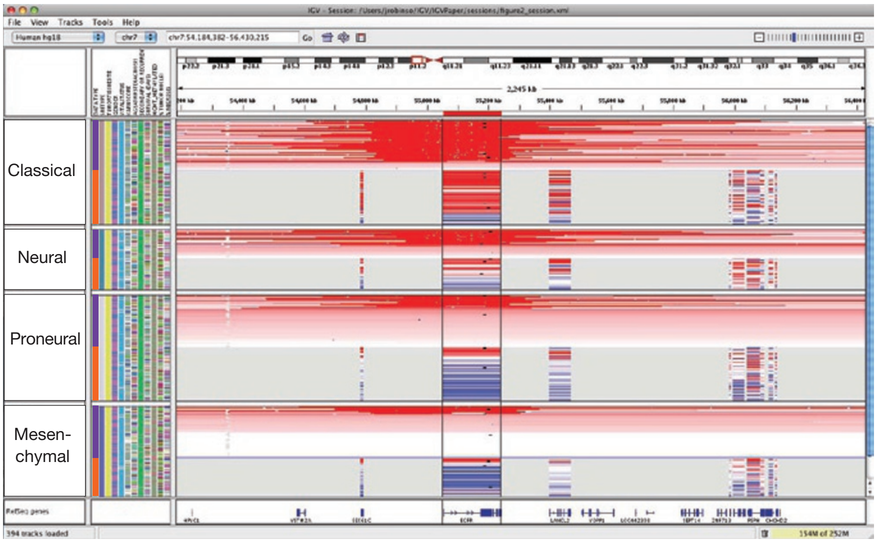
IGV - Session: /Users/jrobinso/IGV/IGVPaper/sessions/figure2_session.xml
File View Tracks Tools Help
Human hg18
chr7
chr7:54,184,382-56,430,215
Go
2,245 kb
Classical
Neural
Proneural
Mesen-
chymal
RefSeq genes
394 tracks loaded
154M of 252M

Figure 1 Copy number, expression and mutation data grouped by tumor subtype. This figure illustrates an integrated, multi-modal view of 202 glioblastoma multiforme samples from The Cancer Genome Atlas (TCGA). Copy number data are segmented values from Affymetrix (Santa Clara, CA, USA) SNP6.0 arrays. Expression data are limited to genes represented on all TCGA-employed platforms and displayed across the entire gene locus. Red shading indicates relative upregulation of a gene and the degree of copy gain of a region; blue shading indicates relative downregulation and copy loss. Small black squares indicate the position of point missense mutations. Samples are grouped by tumor subtype (2nd annotation column) and data type (1st sample annotation column) and sorted by copy number of the EGFR locus. Linking by sample attributes ensures that the order of sample tracks is consistent across data types within their respective tumor subtypes.

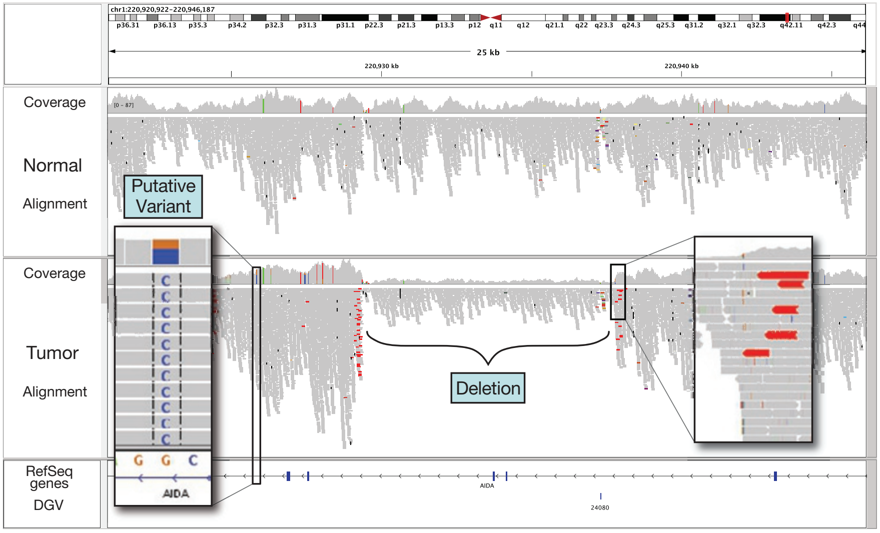

Figure 2 View of aligned reads at 20-kb resolution. Coverage plot and alignments from paired-end reads for a matched tumor/normal pair. Sequencing was performed on an Illumina (San Diego, CA) GA2 platform and aligned with Maq (http://maq.sourceforge.net/). Alignments are represented as gray polygons with reads mismatching the reference indicated by color. Loci with a large percentage of mismatches relative to the reference are flagged in the coverage plot as color-coded bars. Alignments with unexpected inferred insert sizes are indicated by color. There is evidence for a ~10-kb deletion (removing two exons of AIDA) in the tumor sample not present in the normal.

and up to thousands, of samples and correlation of these integrated data sets with clinical and phenotypic variables. A researcher can define arbitrary sample annotations and associate them with data tracks using a simple tab-delimited file format (**Supplementary Notes**). These might include, for example, sample identifier (used to link different types of data for the same patient or tissue sample), phenotype, outcome, cluster membership or any other clinical or experimental label. Annotations are displayed as a heatmap, but more importantly are used for grouping, sorting, filtering and overlaying diverse data types to yield a comprehensive picture of the integrated data set. This is illustrated in **Figure 1**, a view of copy number, expression, mutation and clinical data from 202 glioblastoma samples from The Cancer Genome Atlas project in a 3-kb region around the epidermal growth factor receptor (*EGFR*) locus[1,4]. The investigator first grouped samples by tumor subtype, then by data type (copy number and expression), and finally sorted them by median copy number over the *EGFR* locus. A shared sample identifier links the copy number and expression tracks, maintaining their relative sort order within the subtypes. Mutation data are overlaid on corresponding copy number and expression tracks, based on shared participant identifier annotations. Several trends in the data stand out, such as a strong correlation between copy number and expression and an overrepresentation of EGFR-amplified samples in the 'Classical' subtype.

IGV's scalable architecture makes it well suited for genome-wide exploration of next-generation sequencing (NGS) data sets, including both basic aligned read data as well as derived results, such as read coverage. NGS data sets can approach terabytes in size, so careful management of data is necessary to conserve computer resources and to prevent information overload. IGV varies the displayed level of detail according to resolution scale. At very wide views, such as the whole genome, IGV represents NGS data by a simple coverage plot. Coverage data are often useful for assessing overall quality and diagnosing technical issues in sequencing runs (**Supplementary Fig. 3**), as well as analysis of ChIP-Seq[5] and RNA-Seq[6] experiments (**Supplementary Figs. 4** and **5**).

As the user zooms below the ~50 kb range, individual aligned reads become visible (**Fig. 2**), and putative SNPs are highlighted as allele counts in the coverage plot. Alignment details for each read are available in popup windows (**Supplementary Figs. 6** and **7**). Zooming in further, individual base mismatches become visible, highlighted by color and intensity according to base call and quality. At this level, the investigator may sort reads by base, quality, strand, sample and other attributes to assess the evidence of a variant. This type of visual inspection can be an efficient and powerful tool for variant call validation, eliminating many false positives and aiding in confirmation of true findings (**Supplementary Figs. 6** and 7).

Many sequencing protocols produce reads from both ends ('paired ends') of genomic fragments of known size distribution. IGV uses this information to color-code paired ends if their insert sizes are larger than expected, fall on different chromosomes or have unexpected pair orientations. Such pairs, when consistent across multiple reads, can be indicative of a genomic rearrangement. When coloring aberrant paired ends, each chromosome is assigned a unique color, so that intra- (same color) and inter- (different color) chromosomal events are readily distinguished (**Fig. 2** and **Supplementary Fig. 8**). We note that misalignments, particularly in repeat regions, can also yield unexpected insert sizes and can be diagnosed with the IGV (**Supplementary Fig. 9**).

There are a number of stand-alone, desktop genome browsers available today[7], including Artemis[8], EagleView[9], MapView[10], Tablet[11], Savant[12], Apollo[13] and the Integrated Genome Browser[14]. Many of them have features that overlap with IGV, particularly for NGS sequence alignment and genome annotation viewing. The Integrated Genome Browser also supports viewing array-based data (**Supplementary Table 1** and **Supplementary Notes**). IGV focuses on the emerging integrative nature of genomic studies, placing equal emphasis on array-based platforms, such as expression and copy-number arrays, NGS, as well as clinical and other sample metadata. Indeed, an important and unique feature of IGV is the ability to view all these different data types together and to use the sample metadata to dynamically group, sort and filter data sets (**Fig. 1**). Another important characteristic of IGV is fast data loading and real-time pan and zoom—at all scales of genome resolution and all data set sizes, including data sets comprising hundreds of samples. Finally, we have placed great emphasis on the ease of installation and use of IGV, with the goal of making both the viewing and sharing of their data accessible to end users who are not informatics specialists.

IGV is open source software and freely available (http://www.broadinstitute.org/igv/), including full documentation on use of the software.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS
Thanks to M.-D. Nazaire, IGV software developer, J. Boehm, M. DePristo, M. Garber, M. Reich and B. Weir, who worked closely with the development team with suggestions for features and usability, and J. Bistline for expert assistance with figures. M. Garber provided source code for multiple alignment tracks. This work was supported by The National Institute of General Medical Sciences (R01GM074024), National Cancer Institute (R21CA135827) and National Human Genome Research Institute (U54HG003067).

AUTHOR CONTRIBUTIONS

J.T.R. and H.T. designed and developed the software; W.W., M.G., E.S.L., G.G. and J.P.M. contributed to the design of the interface and data views; J.P.M. and G.G. oversaw the project; and J.T.R., H.T., W.W., G.G. and J.P.M. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

James T Robinson[1], Helga Thorvaldsdóttir[1], Wendy Winckler[1], Mitchell Guttman[1,2], Eric S Lander[1–3], Gad Getz[1] & Jill P Mesirov[1]

[1]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA.
e-mail: mesirov@broad.mit.edu; jrobinso@broadinstitute.org

1. Cancer Genome Atlas Research Network. *Nature* **455**, 1061–1068 (2008).
2. Durbin, R.M. *et al. Nature* **467**, 1061–1073 (2010).
3. The ENCODE Project Consortium. *Science* **306**, 636–640 (2004).
4. Verhaak, R.G. *et al. Cancer Cell* **17**, 98–110 (2010).
5. Guttman, M. *et al. Nature* **458**, 223–227 (2009).
6. Berger, M.F. *et al. Genome Res.* **20**, 413–427 (2010).
7. Nielsen, C., Cantor, M., Dubchak, I., Gordon, D. & Wang, T. *Nat. Methods* **7**, S5–S15 (2010).
8. Rutherford, K. *et al. Bioinformatics* **16**, 944–945 (2000).
9. Huang, W. & Marth, G. *Genome Res.* **18**, 1538–1543 (2008).
10. Bao, H. *et al. Bioinformatics* **25**, 1554–1555 (2009).
11. Milne, I. *et al. Bioinformatics* **26**, 401–402 (2010).
12. Fiume, M., Williams, V., Brook, A. & Brudno, M. *Bioinformatics* **26**, 1938–1944 (2010).
13. Lewis, S.E. *et al. Genome Biol.* **3**, RESEARCH0082.1–0082.14 (2002).
14. Nicol, J.W., Helt, G.A., Blanchard, S.G. Jr., Raja, A. & Loraine, A.E. *Bioinformatics* **25**, 2730–2731 (2009).