Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells
Qiaolin Deng *et al.*
Science **343**, 193 (2014);
DOI: 10.1126/science.1245316

activating (fig. S10), similar to the previously identified *MTOR* Ser^{2215} → Tyr (*37*). Moreover, adjacent regions of this recurrence showed heterogeneous mTOR complex 1 (mTORC1) activity (Fig. 3E and fig. S11). Microdissection revealed that although these adjacent regions shared a subset of the mutations found in the initial tumor, *MTOR* S2215F and other TMZ-associated mutations were present only in the region that stained strongly for mTORC1 activation, which also had higher staining of the proliferation marker Ki-67, implying that the TMZ-associated mutations conferred a proliferative advantage. A distal second recurrence harbored the same TMZ-associated mutations and stained strongly and homogeneously for mTORC1 targets (fig. S12). Although both regions of the first recurrence were GBM, the hypermutated subclone underwent in vivo selection, invaded distally, and seeded the second recurrence (figs. S13 and S14). Across our cohort, Akt-mTOR pathway mutations corresponded with elevated phospho-4E-BP1 and RPS6 in vivo, indicating hyperactivated mTORC1 in recurrent GBMs relative to their initial tumors (fig. S12).

There was no evidence that the mutations in the RB and Akt-mTOR signaling pathways preceded TMZ treatment, according to analysis of additional geographically distinct samples of initial tumors from four of the six patients with hypermutated recurrent tumors (table S7). Non-hypermutated recurrent tumors that progressed to GBM also acquired genetic changes in these signaling pathways, but through alternative mechanisms. In contrast, none of the grade II-III recurrences acquired mutations in these pathways. These data suggest a connection among TMZ treatment, driver mutations in oncogenic signaling pathways, and malignant progression.

Through direct comparison of the genomic landscape of gliomas at initial diagnosis and recurrence, we were able to infer the mutational character of the infiltrating tumor cells that give rise to recurrence and that adjuvant therapy with TMZ is intended to eliminate. Recurrences did not typically arise from cells bearing the full set of mutations found in the initial tumor, as would be expected from a local recurrence in the absence of selective pressure from adjuvant chemotherapy. This finding complicates the use of tumor genomics to design precision therapies targeting residual disease. We also demonstrated an alternative evolutionary path of low-grade glioma that is largely determined by adjuvant chemotherapy with TMZ. This extends earlier studies of primary GBMs (*23*, *25*), unpaired recurrent tumors (*22*), and a cell culture model (*20*). Future basic and clinical studies must weigh the initial antitumor effects of TMZ against the potential risk of inducing new driver mutations and malignant progression. Ultimately, a better understanding of the invading cells that give rise to recurrent tumors and the effect of adjuvant therapeutics on their evolution will facilitate the development of new strategies to delay or prevent recurrence and malignant progression.

References and Notes

1. M. Gerlinger, C. Swanton, *Br. J. Cancer* **103**, 1139–1143 (2010).
2. M. Greaves, C. C. Maley, *Nature* **481**, 306–313 (2012).
3. D. A. Landau *et al.*, *Cell* **152**, 714–726 (2013).
4. L. Ding *et al.*, *Nature* **481**, 506–510 (2012).
5. M. Gerlinger *et al.*, *N. Engl. J. Med.* **366**, 883–892 (2012).
6. S. Yachida *et al.*, *Nature* **467**, 1114–1117 (2010).
7. X. Wu *et al.*, *Nature* **482**, 529–533 (2012).
8. N. Sanai, S. Chang, M. S. Berger, *J. Neurosurg.* **115**, 948–965 (2011).
9. M. Westphal, K. Lamszus, *Nat. Rev. Neurosci.* **12**, 495–508 (2011).
10. See supplementary materials on *Science* Online.
11. Y. Jiao *et al.*, *Oncotarget* **3**, 709–722 (2012).
12. T. Watanabe, S. Nobusawa, P. Kleihues, H. Ohgaki, *Am. J. Pathol.* **174**, 1149–1153 (2009).
13. K. Watanabe *et al.*, *Clin. Cancer Res.* **3**, 523–530 (1997).
14. P. P. Medina *et al.*, *Hum. Mutat.* **29**, 617–622 (2008).
15. S. Glaros, G. M. Cirrincione, A. Palanca, D. Metzger, D. Reisman, *Cancer Res.* **68**, 3689–3696 (2008).
16. D. Rohle *et al.*, *Science* **340**, 626–630 (2013).
17. N. J. Szerlip *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3041–3046 (2012).
18. A. Sottoriva *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4009–4014 (2013).
19. L. M. F. Merlo, J. W. Pepper, B. J. Reid, C. C. Maley, *Nat. Rev. Cancer* **6**, 924–935 (2006).
20. W. J. Bodell, N. W. Gaikwad, D. Miller, M. S. Berger, *Cancer Epidemiol. Biomarkers Prev.* **12**, 545–551 (2003).
21. C. Greenman *et al.*, *Nature* **446**, 153–158 (2007).
22. Cancer Genome Atlas Research Network, *Nature* **455**, 1061–1068 (2008).
23. C. Hunter *et al.*, *Cancer Res.* **66**, 3987–3991 (2006).
24. D. P. Cahill *et al.*, *Clin. Cancer Res.* **13**, 2038–2045 (2007).
25. S. Yip *et al.*, *Clin. Cancer Res.* **15**, 4622–4629 (2009).
26. G. Reifenberger, J. Reifenberger, K. Ichimura, P. S. Meltzer, V. P. Collins, *Cancer Res.* **54**, 4299–4303 (1994).
27. H. Wang *et al.*, *Lab. Invest.* **84**, 941–951 (2004).
28. D. N. Louis, *Annu. Rev. Pathol.* **1**, 97–117 (2006).
29. T. Tsai *et al.*, *Arch. Ophthalmol.* **122**, 239–248 (2004).
30. C. Houdayer *et al.*, *Hum. Mutat.* **23**, 193–202 (2004).
31. X. Q. Qin, T. Chittenden, D. M. Livingston, W. G. Kaelin Jr., *Genes Dev.* **6**, 953–964 (1992).
32. J. Koh, G. H. Enders, B. D. Dynlacht, E. Harlow, *Nature* **375**, 506–510 (1995).
33. S. A. Forbes *et al.*, *Curr. Protoc. Hum. Genet.* Chapter 10, Unit 10.11 (2008).
34. M. C. Fargnoli *et al.*, *J. Invest. Dermatol.* **111**, 1202–1206 (1998).
35. S. Kang, A. G. Bader, P. K. Vogt, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 802–807 (2005).
36. S. Y. Han *et al.*, *Cancer Res.* **60**, 3147–3151 (2000).
37. T. Sato, A. Nakashima, L. Guo, K. Coffman, F. Tamanoi, *Oncogene* **29**, 2746–2752 (2010).

Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells

Qiaolin Deng,[1]* Daniel Ramsköld,[1,2]* Björn Reinius,[1,2] Rickard Sandberg[1,2]†

Expression from both alleles is generally observed in analyses of diploid cell populations, but studies addressing allelic expression patterns genome-wide in single cells are lacking. Here, we present global analyses of allelic expression across individual cells of mouse preimplantation embryos of mixed background (CAST/EiJ × C57BL/6J). We discovered abundant (12 to 24%) monoallelic expression of autosomal genes and that expression of the two alleles occurs independently. The monoallelic expression appeared random and dynamic because there was considerable variation among closely related embryonic cells. Similar patterns of monoallelic expression were observed in mature cells. Our allelic expression analysis also demonstrates the de novo inactivation of the paternal X chromosome. We conclude that independent and stochastic allelic transcription generates abundant random monoallelic expression in the mammalian cell.

In diploid organisms, the zygote inherits one set of autosomal chromosomes from each parent. Although it is widely believed that transcription of autosomal genes occurs from both parental alleles, specific classes of genes have been shown to express only one, randomly

selected, allele (allelic exclusion) (1–3). Analyses of clonally amplified lymphocytes by using single-nucleotide polymorphism (SNP)–sensitive microarrays revealed that 8% of human autosomal genes and 16% of mouse genes showed a type of random, monoallelic expression that was stably maintained during clonal expansion (4, 5). Furthermore, parental-specific (imprinted) expression has been demonstrated for 1% of autosomal genes (6, 7) and, perhaps most strikingly, in the inactivation of one X chromosome in female cells (8). Although RNA fluorescent in situ hybridization (RNA-FISH) has been used to study a few individual genes (9, 10), little is known about general patterns of allelic expression in single cells.

To investigate allele-specific gene expression at single-cell resolution, we isolated 269 individual cells dissociated from in vivo F1 embryos (CAST/EiJ × C57BL/6J, hereafter abbreviated as CAST and C57, respectively) from oocyte to blastocyst stages of mouse preimplantation development (PD) (11). We generated transcriptome profiles with Smart-seq (12) or Smart-seq2 (13) from each individual cell (table S1 and fig. S1, A and B). Principal component analysis (PCA) clustered the cells by developmental stage and embryo, effectively reconstructing the dynamics of PD (Fig. 1A). Next, using strain-specific SNPs (14) to distinguish transcription from the maternal and paternal chromosomes (15) we observed that 82% of all genes expressed during PD contained ≥1 informative SNP (fig. S1C) and that different SNPs within the same gene gave coherent allelic calls (fig. S2). Because maternal RNA lingers from the oocyte (16), we expected the maternal genotype to dominate the zygotic

[1]Ludwig Institute for Cancer Research, Box 240, 171 77 Stockholm, Sweden. [2]Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden.

*These authors contributed equally to this work.
†Corresponding author. E-mail: rickard.sandberg@ki.se

transcriptome. Indeed, the zygote, and also the early two-cell, contained essentially only maternal RNA, but in the subsequent stages, the maternal fraction gradually declined to reach parity with paternal transcripts at the four-cell stage (Fig. 1B), which is consistent with rapid maternal transcript clearance and zygotic genome activation. As a control for the accuracy in alignments and SNP annotation, we analyzed individual cells of pure C57 or CAST background and found 99.4 and 99.7% correctly classified reads, respectively (Fig. 1B).

We next investigated the gene activation across paternal chromosomes. We found that genes on the paternal X chromosome (Xp) of female embryos were indeed transcriptionally activated similarly to those on paternal autosomes during a defined time window of the PD. Subsequent reinactivation occurred first beyond the four-cell stage (Fig. 2A), demonstrating de novo Xp inactivation (17, 18) rather than inheritance and propagation of a pre-inactivated Xp (19). X chromosome inactivation initiates from the X-inactivation center (Xic), from which Xist is transcribed, and spreads in cis (18). Our data provided a high-resolution map of silencing over Xp at the four-cell, 16-cell, and early blastocyst stages (Fig. 2B) that substantiates the observation of a silencing gradient (19) and demonstrated that the spread of Xp silencing is not a simple function of the distance to Xic (Fig. 2B, and escapee genes in fig. S3 and table S2).

These findings gave us confidence in inferring biological signals from the allelic information in single-cell RNA sequencing (RNA-seq). To further explore allelic expression of autosomal genes, we classified their expression as biallelic, maternal monoallelic, or paternal monoallelic according to SNP-containing reads. Surprisingly, this revealed a great degree of monoallelic expression (on average, 54% of genes) across all stages of PD (figs. S4 to S24). The monoallelic

calls were similar or more abundant than in available RNA-FISH data (fig. S25) (9, 10). Because single-cell transcriptome methods suffer from stochastic losses of RNA species, it was necessary to determine to what extent random sampling effects inflate observed monoallelic calls. We therefore lysed individual cells (from 8- or 16-cell embryos) and split the lysate into two equal volume fractions that were independently processed into sequencing libraries. Using the allelic calls from the split-pairs, we modeled the stochastic losses and inferred the underlying levels of biallelic and monoallelic expression in sets of genes binned by expression level (figs. S26 and S27). We estimated that 60% of all polyadenylated [poly (A)$^+$] RNA molecules are lost in the Smart-seq2 protocol (Fig. 3A) (13) because inferred losses stabilized at levels equal to a single RNA molecule. This analysis uncovered coherent monoallelic expression estimates across independent split-cell experiments with a median of 17% of genes (Fig. 3B). Although technical losses of RNA contributed as much as 66% of observed monoallelic expression, this strategy allowed us to determine the underlying amount of biological monoallelic expression in single cells.
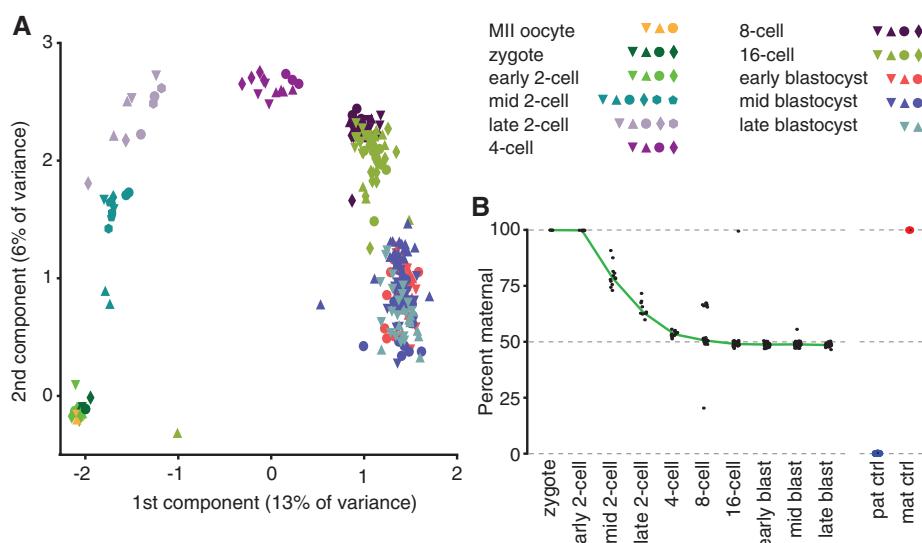
In subsequent analyses of monoallelic expression, we focused only on transcripts expressed at sufficient abundances to be little influenced by random sampling, as determined in our control experiments (fig. S28). Exploring the levels of monoallelic expression among cells from the four-cell stage to the late blastocyst stage, we observed similar levels throughout the PD, with an average of 12 to 24% monoallelic expression for mRNAs (Fig. 3D) and 19 to 26% for noncoding poly(A)$^+$ RNAs. In contrast, consistent biallelic expression was observed for only a few hundred genes (table S3), often with housekeeping functions (4, 9). Allele classification of single-cell data of pure C57 and CAST background gave 97.3 and 99.5% correct monoallelic calls, respectively (Fig. 3D).

Fig. 1. Single-cell transcriptomes reconstruct preimplantation development. (A) Single-cell gene expression profiles projected onto the first two principal components. Cells from different stages and embryos are designated by colors and symbols. (B) The percentage (by mass) of maternal RNA observed in single-cell transcriptomes (black dots; median is in green) at different stages of development and in controls from pure maternal (CAST, red) and paternal (C57, blue) backgrounds. The eight-cell stage outlier cells with maternal bias are all from one embryo (supplementary text).
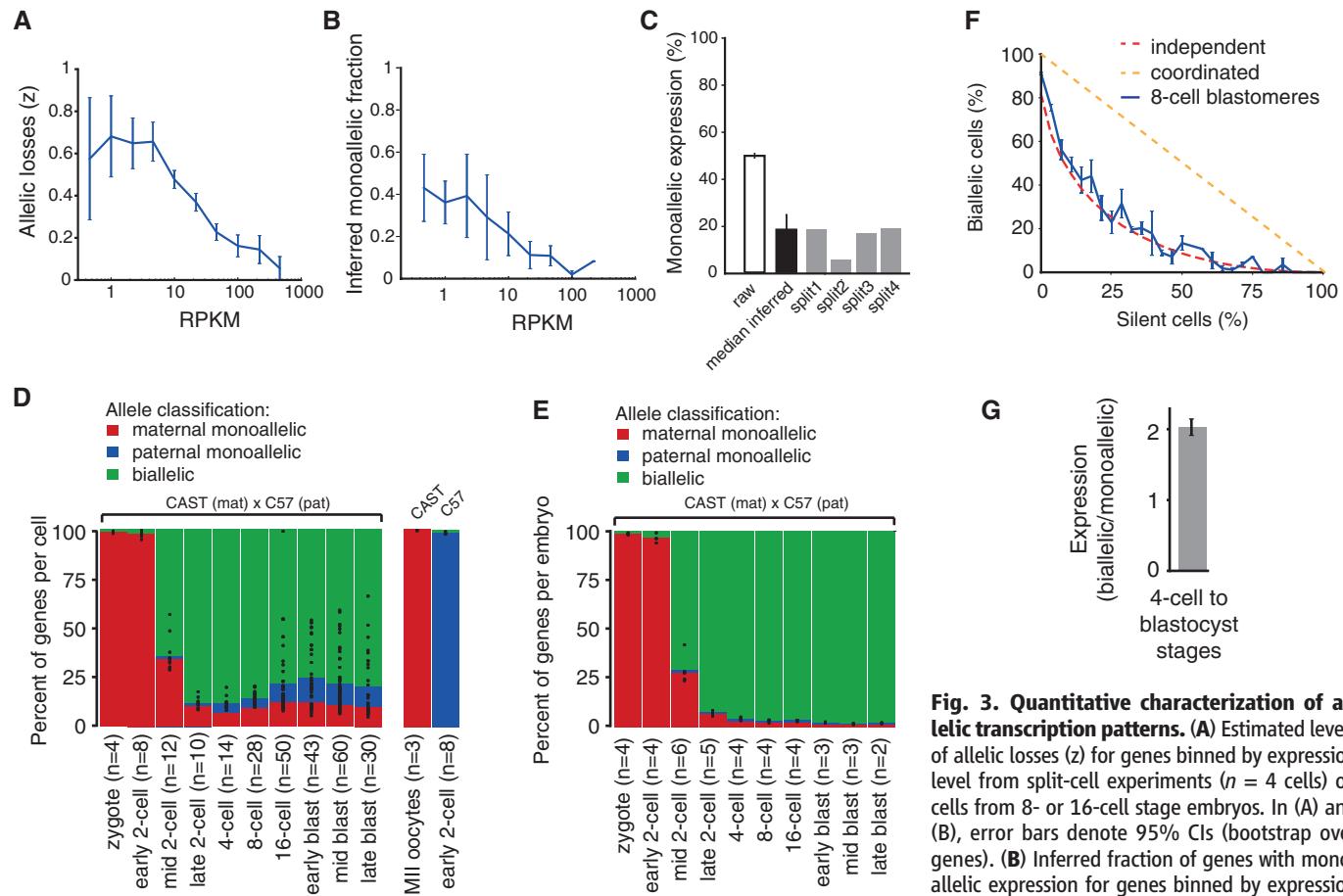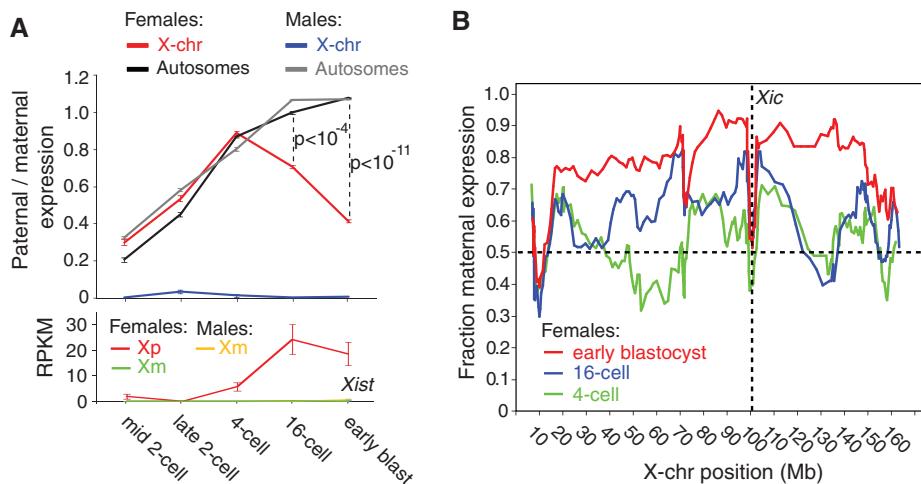
Pooling cells by embryo removed essentially all monoallelic expression (Fig. 3E), demonstrating a high degree of cell-specific randomness in mono-allelic expressions. We therefore concluded that a dynamic type of random monoallelic expression is abundant in blastomeres.
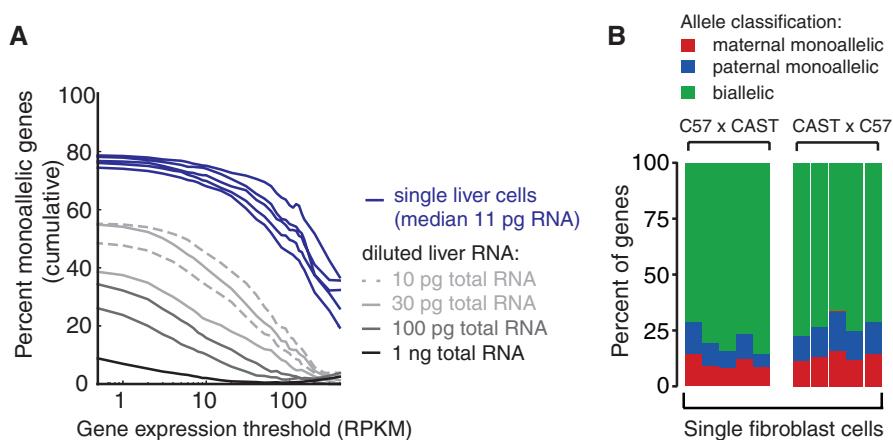
Pioneer studies on interleukin-4 (20, 21) found transcription of the two alleles to occur independently, but it remained unknown whether such

Fig. 2. De novo paternal X-chromosome inactivation. (**A**) (Top) Average ratio of SNP-containing reads (paternal, C57; maternal, CAST) per autosome or X chromosome in cells of embryos (male and female, separately) at stages of preimplantation development. Error bars denote SEM (n = 3 to 28 cells), and P values were calculated using Student's t-test. (Bottom) Allele-resolution gene expression (RPKM) of *Xist* in cells from different developmental stages. Colors indicate embryos according to sex and allele-specific expression. Only stages for which female embryos were available are shown. Error bars denote SEM (n = 3 to 28 cells). (**B**) Spread of Xp inactivation, shown as the fraction maternal reads for genes on the X chromosome by using a moving window average of 10 genes. *Xic* is marked by a dashed line.

Percent of genes with monoallelic expression in 8- or 16-cell stage embryo cells, from raw observations, individual split-cell experiments, and their median. Error bars denote 95% CIs. (**D**) Mean percent of genes with monoallelic expression across cells at each developmental stage for genes with mean expression ≥90 RPKM per stage (fig. S28). Individual cell percentages are shown as black dots. (**E**) Mean percent of genes with monoallelic expression per embryonic stage, when pooling cells from the same embryo stage. Individual embryo percentages are shown as black dots. Colors and RPKM threshold are as in (D). (**F**) Mean fraction of cells (eight-cell stage) with biallelic expression for genes binned by their fraction of nondetected (silent) cells; error bars give SEM (n ≥ 3 genes). The expected biallelic fraction was computed on the basis of either a model of independent allele activation or coordinated activation (11). (**G**) Mean ratio of total expression between cells with biallelic or monoallelic expression of the same genes. Error bars give SEM (n = 120 genes).

Fig. 3. Quantitative characterization of allelic transcription patterns. (**A**) Estimated levels of allelic losses (z) for genes binned by expression level from split-cell experiments (n = 4 cells) on cells from 8- or 16-cell stage embryos. In (A) and (B), error bars denote 95% CIs (bootstrap over genes). (**B**) Inferred fraction of genes with monoallelic expression for genes binned by expression level, using split-cell experiments as in (A). (**C**)

Fig. 4. Monoallelic expression in different mature cell types. (**A**) The cumulative fraction of monoallelic expression as a function of the expression levels for individual liver cells and at different dilutions. (**B**) Percentage of genes with monoallelic expression in individual cultured adult fibroblasts (C57 × CAST and CAST × C57) for genes with mean expression ≥20 RPKM per stage (fig. S28).

independence applies on the genomic scale. Using our comprehensive allelic expression data for thousands of genes, we therefore investigated whether expression from the two alleles occurred independently of one another. Independent allelic expression would yield a specific relationship between the fraction of cells with biallelic, monoallelic, or no expression (10), which is different from scenarios of coordinated allelic expression. Markedly, the observed fraction of cells with biallelic expression followed the fraction expected under the independence model (Fig. 3F, eight-cell stage data, and fig. S29, all stages), demonstrating independent allelic transcription for genes across all expression levels, substantially extending results on individual genes (20, 21) to a global principle. Under independent allelic expression, biallelic expression of a particular gene would on average result in twofold higher RNA copy numbers than would the same gene in cells in which the gene is monoallelically expressed because cells with biallelic expression have transcriptional output from two alleles. To test this hypothesis, we analyzed mean gene expression levels in cells with biallelic expression and indeed observed them to be 2.0 ± 0.1 times higher [95% confidence interval (CI), bootstrap] than the levels in cells with monoallelic expression at all developmental stages (Fig. 3G). Thus, both the allelic expression patterns and the expression levels point to independent allelic transcription.

Because embryonic cells are uncommitted progenitors, it was important to determine whether abundant random monoallelic expression also occurs in mature cells. To this end, we investigated single-cell transcriptomes of in vivo liver cells (C57 × CAST), together with a control dilution series of RNA extracted from liver tissue (C57 × CAST) (Fig. 4A and fig. S30). The monoallelic expression levels in single liver cells were higher than sampling effects measured in control dilutions across all expression levels (Fig. 4B), and the fraction of true monoallelic calls increased with the expression threshold. We also profiled 10 individual adult mouse fibroblasts (five from each reciprocal cross) and detected similar random monoallelic expression, on average 24% of

expressed genes per cell. We therefore conclude that random monoallelic expression is abundant in both embryonic and mature cells.

In this study, we uncovered a stochastic pattern of monoallelic expression that differs from the stable allelic regulation of genomic imprinting and allelic exclusion (22, 23). It also differs from the stably maintained monoallelic expression observed in clonal lymphoid cell populations (4, 5). Instead, the rapid expression dynamics that we uncovered in individual cells are consistent with models of transcriptional bursting (24). In each cell, independent bursts of transcription occur from both alleles over time, but RNA from only one allele is often present at any given time. Because stochastic losses of RNA substantially inflates naive estimates of allelic expressions, stringent controls such as split-cells and dilution series are of critical importance for accurate allelic expression analyses in single cells. It is likely that stochastic transcription of heterozygous alleles contributes to variable expressivity—phenotypic variation among cells and individuals of identical genotypes—which may have fundamental implications for variable disease penetrance and severity (25–28).

References and Notes

1. B. Pernis, G. Chiappino, A. S. Kelus, P. G. Gell, *J. Exp. Med.* **122**, 853–876 (1965).
2. M. Malissen et al., *Immunol. Today* **13**, 315–322 (1992).
3. A. Chess, I. Simon, H. Cedar, R. Axel, *Cell* **78**, 823–834 (1994).
4. A. Gimelbrant, J. N. Hutchinson, B. R. Thompson, A. Chess, *Science* **318**, 1136–1140 (2007).
5. L. M. Zwemer et al., *Genome Biol.* **13**, R10 (2012).
6. W. Reik, J. Walter, *Nat. Rev. Genet.* **2**, 21–32 (2001).
7. B. DeVeale, D. van der Kooy, T. Babak, *PLOS Genet.* **8**, e1002600 (2012).
8. J. T. Lee, *Nat. Rev. Mol. Cell Biol.* **12**, 815–826 (2011).
9. Y. Miyanari, M.-E. Torres-Padilla, *Nature* **483**, 470–473 (2012).
10. S. Kalantry, S. Purushothaman, R. B. Bowen, J. Starmer, T. Magnuson, *Nature* **460**, 647–651 (2009).
11. Materials and methods are available as supplementary materials on *Science* Online
12. D. Ramsköld et al., *Nat. Biotechnol.* **30**, 777–782 (2012).
13. S. Picelli et al., *Nat. Methods* **10**, 1096–1098 (2013).
14. T. M. Keane et al., *Nature* **477**, 289–294 (2011).
15. T. Babak et al., *Curr. Biol.* **18**, 1735–1741 (2008).
16. L. Pikó, K. B. Clegg, *Dev. Biol.* **89**, 362–378 (1982).
17. I. Okamoto, A. P. Otte, C. D. Allis, D. Reinberg, E. Heard, *Science* **303**, 644–649 (2004).
18. I. Okamoto et al., *Nature* **438**, 369–373 (2005).
19. K. D. Huynh, J. T. Lee, *Nature* **426**, 857–862 (2003).
20. I. Rivière, M. J. Sunshine, D. R. Littman, *Immunity* **9**, 217–228 (1998).
21. M. Bix, R. M. Locksley, *Science* **281**, 1352–1354 (1998).
22. S. Feng, S. E. Jacobsen, W. Reik, *Science* **330**, 622–627 (2010).
23. P. K. Yang, M. I. Kuroda, *Cell* **128**, 777–786 (2007).
24. A. Raj, A. van Oudenaarden, *Cell* **135**, 216–226 (2008).
25. M. Muenke et al., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8102–8106 (1994).
26. P. Janku et al., *Am. J. Med. Genet.* **5**, 117–123 (1980).
27. J. C. Carey, D. H. Viskochil, *Am. J. Med. Genet.* **89**, 7–13 (1999).
28. A. Raj, S. A. Rifkin, E. Andersen, A. van Oudenaarden, *Nature* **463**, 913–918 (2010).