DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Reading and writing omes

Citation	Church, George M. 2013. Reading and writing omes. Molecular Systems Biology 9(1): 642.
Published Version	doi:10.1038/msb.2012.75
Accessed	February 19, 2015 11:59:59 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:10588028
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

molecular
systems
biology

EDITORIAL

Reading and writing omes

'Systems Technologies' are increasingly potent drivers of biological research. *Molecular Systems Biology* will be illustrating this evolution with a new Reviews Series highlighting key technologies in systems medicine, genome-scale, computational, quantitative and synthetic biology. The series is launched with a review from the Snyder group on reading human omes (Soon *et al*, 2013) and a companion review on writing genomes from Harvard's Wyss Institute (Esvelt and Wang, 2013).

Past achievements, future milestones

Exponential improvement in reading and writing technologies (Carr and Church, 2009) (1.5-fold/year since 1960s, 6-fold per year since 2005) created a series of breakthroughs: The first genome read was MS2 in 1976 (phiX in 1977); first written was hepatitis C virus in 2000 (Blight *et al*, 2000) (polio in 2002). The first bacterial genome read was *Helicobacter* in 1994 (*Haemophilus* in 1995). The first genome transplanted from *in vitro* DNA into radically foreign cytoplasm was *Synechocystis* into *Bacillus* in 2005 (then *Mycoplasma mycoides* into similar cytoplasm in 2007). Significantly, so far, no vertebrate genome has been fully read, due to repetitive regions, and no new organism function has been achieved by genome-scale writing. We expect to see breakthroughs on both fronts in 2013.

Utility beyond research feeding more research

The first widely used genetic engineering vector was pBR322, constructed and sequenced 1977-1978, parts of which are still present in modern vectors. This enabled dissection of previously recalcitrant biological systems via pure components and swiftly lead to commercial production of a stream of human proteins, including insulin, interferons, epo and therapeutic antibodies. Proteins still constitute the fastest growing category of new therapeutics. Comparative genomics, metabolic engineering and systems biology (Schirmer *et al*, 2010) have resulted in factories already at production-scale for chemicals, fuels and pharmaceuticals.

What next for reading omes?

Some say that due to other costs, the plummeting human genome price will stop at $1000, but the million-fold cost improvement changes not only how we read our once-in-a-lifetime inherited genomes, but also how we can measure our day-to-day immunome response to our microbiome, cancer transcriptome and allergome. Portable nanopore devices with minimal sample handling and 100-kbp reads could enable real-time environmental air and food monitoring. Nanotags seem ready to greatly improve raw sequencing accuracy (Kumar *et al*, 2012) and detection of modified bases (Korlach and Turner, 2012). Another technology with a potentially huge impact on systems biology will be fluorescent *in situ* sequencing, enabling studies of not just single cells, but subcellular and multicellular features, and reveal tumor and developmental cell-to-cell heterogeneity. Combined with super-resolution fluorescent microscopy, *in situ* measures will reveal the 3D structure of genomes, epigenomes and cells (Beliveau *et al*, 2012). This allows us to go beyond ENCODE (Bernstein *et al*, 2012) and 'Organs-on-chips', which pragmatically employ cancer-like cells and primary cells from poorly documented human sources, to well-defined open-access personal genome cells (Ball *et al*, 2012), engineered human cells and even human plus bacterial cells in synthetic gut ecosystems (Kim *et al*, 2012).

From reading to diagnostics

Will we need $100 000 to interpret our $1000 genome? Certainly not. Automation of data analysis workflows and minimization of false-positive diagnostic outcomes already deliver full genome interpretation for $400 per genome (a fraction of the $4000 raw sequence). Interpretation will expand from simple Mendelian models to multigenic, multi-environmental component systems models. This transition will benefit from shareable integrated 'precision medicine' data sets on individuals (not averages). Despite the increase in actionable gene tests from a few in 1990–2700 today, a vocal few insist that personal genomics is not worth it. Yes; DNA like many other diagnostics may not reveal anything new, but you don't know until you look. We should not restrict gene tests by family history, as many afflicted are the first in their family. Notably as we learn to better control environmental factors, the genetic component (heritability) of a disease can increase (unbounded by previous association studies or twin studies conducted before the reduction in environmental components). Ironically, the push for larger cohorts in the name of statistical power, results in confounding via lumping of disparate types. Focusing instead on phenotypic extremes (positive and negative) can result in clearer diagnostics, preventatives and therapies applicable across the whole spectrum. Furthermore, systems approaches focused on causality rather than correlation seem quite promising, even with cohorts as small as $N = 1$. Examples of going from genome-wide analyses to treatment are accumulating (e.g., Nic Volker, the Beery twins, Mike Snyder, John Lauerman).
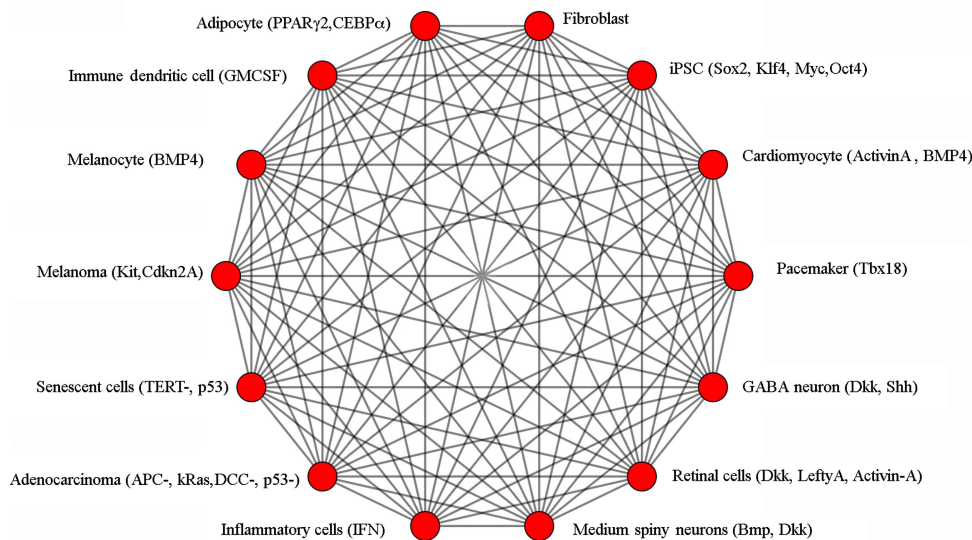
Figure 1 The epigenomic transition matrix. New systems technologies will allow reading diverse epigenomic states *in situ* and writing (reprogramming) these in recipient systems for personal diagnostic, transplantation and other therapeutic purposes.

What next for writing omes? Genomic and epigenomic grand challenges

The first question is why? Why genome-wide rather than a few genes? Genome engineering enables non-standard amino acids, safety isolation and multi-virus resistance (Isaacs *et al*, 2011). Making one genome at a time at high cost (albeit decreasing) misses a key advantage not available to other engineering fields, which is the ability to use system knowledge and clever selections on billions of genomes. Construction of such billions benefits from synthesis of raw oligos on chips and using combinatorial multiplex automated genome engineering. For more difficult organisms, we need extra guidance for genomic and epigenomic reprogramming via Zn-Fingers, TALEs or CRISPR (Mali *et al*, 2012). We will see growing use of sequencing to quantitate phenotypes of large libraries (of codons, *cis*-regulatory signals, etc) and library-by-library measures (antibodies versus antigens, RNA versus protein, etc). The ability to synthesize and deliver complex mixtures (Kim *et al*, 2011) of mRNAs, miRNAs, siRNAs and gRNAs put us on the verge of a transition matrix among all normal and pathological epigenomic states, and therapies (Figure 1).

From writing to therapeutics

Just as proteins are 'smarter' than small molecules, cells are smarter still. Genetic therapy has transitioned from random viral payload integration in the 1990 s to precise targeting today. A Zn-finger-nuclease targeting CCR5 DNA is a promising treatment for AIDS now in phase 2 trials. Also remarkable is the extension of the concept of enhancing drugs and devices—such as cognitive enhancers—to the notion of enhancing gene therapies that will 'cure' people of their common genotype using a minor variant—rather than the older goal of fixing rare genetic diseases using the common variant.

In conclusion, reading and writing technologies are now extending across a broad range of physical and multiplexing scales. Combining multiplexing at the sequence level with parallelized sample processing provide biologists with system-wide functional testing approaching with sufficient power to match the large-scale hypothesis generation that typically results from ome data.

Conflict of interest

The author declares that he has no conflict of interest.

George M Church
Department of Genetics, Harvard Medical School, Boston, MA, USA

References

Ball MP, Thakuria JV, Zaranek AW, Clegg T, Rosenbaum AM, Wu X, Angrist M, Bhak J, Bobe J, Callow MJ, Cano C, Chou MF, Chung WK, Douglas SM, Estep PW, Gore A, Hulick P, Labarga A, Lee JH, Lunshof JE *et al* (2012) A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci USA* **109:** 11920–11927

Beliveau BJ, Joyce EF, Apostolopoulos N, Yilmaz F, Fonseka CY, McCole RB, Yiming Chang Y, Li JB, Senaratne TN, Williams BR, Rouillard JM, Wu CT (2012) A versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc Natl Acad Sci USA* **109:** 21301–21306(PNAS; PMID: 23236188)

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, *et al* for The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74

Blight KJ, Kolykhalov AA, Rice CM (2000) Efficient initiation of HCV RNA replication in cell culture. *Science* **290:** 1972–1974

Carr P, Church GM (2009) Genome engineering. *Nature Biotech* **27:** 1151–1162

Esvelt KM, Wang HW (2013) Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol* **9:** 641

Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen A, Gianoulis T, Goodman D, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333:** 348–353

Kim HJ, Huh D, Hamilton G, Ingber DE (2012) Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. *Lab Chip* **12:** 2165–2174

Kim TK, Sul JY, Peternko NB, Lee JH, Lee M, Patel VV, Kim J, Eberwine JH (2011) Transcriptome transfer provides a model for understanding the phenotype of cardiomyocytes. *Proc Natl Acad Sci USA* **108:** 11918–11923

Korlach J, Turner SW (2012) Going beyond five bases in DNA sequencing. *Curr Opin Struct Biol* **22:** 251–261

Kumar S, Tao C, Chien M, Hellner B, Balijepalli A, Robertson JW, Li Z, Russo JJ, Reiner JE, Kasianowicz JJ, Ju J (2012) PEG-Labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci Rep* **2:** 684

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville J, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* (e-pub ahead of print 3 January 2013; doi:10.1126/science.1232033)

Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* **329:** 559–562

Soon WW, Hariharan M, Snyder MP (2013) High-throughput sequencing for biology and medicine. *Mol Syst Biol* **9:** 640