

## EDITORIAL

# A decade and genome of change

Clare Garvey, Editor, *Genome Biology*\*

### Abstract

Innovations in genomic technologies have generated huge advances in biomedical research over the last decade.

Ten years ago, Bill Clinton, the then US president, announced at an historic event at the White House that the international Human Genome Project and Celera Genomics corporation had completed the initial draft of the human genome. President Clinton pledged the US's commitment to continue to translate this genomic advance into healthcare and therapeutic strategies, as well as protecting private genetic information. Little did he know of the impending cuts in exactly this area during the Bush years - ah, but let's not go there. The genomics field since then has progressed at a phenomenal rate, with advances in the field being nothing short of monumental.

At around the same time as this historic announcement, *Genome Biology* [1] was launched. This new journal was quite unlike other journals in that it was open access and published online. In an accompanying column to this editorial, Greg Petsko [2], *Genome Biology's* long-term and beloved-of-many columnist, discusses *Genome Biology's* launch, in addition to charting our success and the unique approach that has seen *Genome Biology*, in a relatively short period of time, take its place as a premier journal for genomics research. To mark some of the developments in the genomics field in the past decade, and to celebrate *Genome Biology's* tenth birthday, we have commissioned a series of reviews focusing on key areas from the last ten years, ranging from the human microbiome to the cancer genome projects. The themes of these reviews will also be discussed at *Genome Biology's* inaugural conference, which is being hosted jointly with our sister journal, *Genome Medicine* [3], in Boston in October [4].

Technological developments over the past decade have been the catalyst of innovation and progress, driving the

genomics field forward at a dizzying pace. Along with these technological advances have come some revisions of the very gene count estimates that were announced ten years ago. Strikingly, current estimates are nowhere near the original 40,000 genes that humans were estimated to have. In his review entitled 'Between a chicken and a grape: estimating the number of human genes' Steven Salzberg [5] reveals that current estimates of the true human gene count are closer to 20,000. So why has the estimated gene count dropped so dramatically? The advent of computational gene prediction and comparative genome mapping methods is largely responsible for the revision and, more recently, technological advances that have allowed small RNAs and alternative splice forms to be identified have also played a part. Salzberg predicts similar gene count estimate revisions for the chicken and grape genomes to those seen for the human genome.

The study of individual genes is now considered to be reductionist though; it is now in vogue to consider a more systems-led approach, whereby all protein-protein interactions are documented and, on top of this, transcriptional, metabolomic and even environmental input is layered, providing a multidimensional readout of a cell's activity. The rise of the systems biology field is outlined in a review by Nevan Krogan and Michael Fischbach [6]. In a similar vein, it is now apparent that we are not simply the sum of our genes. The contribution to our development and immunity from bacterial communities that reside in the human body is relatively unknown. The Human Microbiome Project (HMP) [7] and the MetaHit consortia [8] aim to categorize these microbial communities and their effect on human health. Rob Knight [9] of the HMP discusses the sequencing depth that is needed to map variation in the human microbial ecosystem between and within individuals.

Various cancer genome projects, such as the Cancer Genome Atlas [10] and the International Cancer Genome Consortium [11], have over the past few years started to reveal mutational signatures for various cancers that will aid targeted treatment and which will provide insights at the stage of diagnosis. In addition, certain cancers, such as acute myeloid leukemia (AML), are being sequenced and compared with the normal genome of that patient to identify disease-causing mutations. These advances have been made mostly through huge developments in

\*Correspondence: Clare.Garvey@genomebiology.com  
Genome Biology, BioMed Central, Gray's Inn Road, London, WC1X 8HL, UK

sequencing technologies. Elaine Mardis [12] discusses the mutations and tumor-specific alterations that have been identified using next generation platforms and which have revolutionized clinical diagnosis of cancer and subsequent therapies. As with all large consortium efforts, vast quantities of sequencing data are generated and, eventually, one has to address the issue of what to do with all of the data. How much should be stored and how can these data be readily accessed in an efficient way, while at the same time protecting genetic privacy, as promised by Clinton ten years ago? Lincoln Stein [13] discusses how cloud computing provides the solution to these hurdles. The doubling time to generate such data now outstrips the rate at which institutes can upgrade data storage facilities. The cost of sequencing has also dropped dramatically compared with the cost of data storage; bizarrely, it is now reasonable to consider re-sequencing a sample instead of storing the raw data from the original read. Cloud computing is the future for bioinformatics analyses: instead of the data user moving the data to the compute cluster, the user moves the compute cluster to the data. Genetic privacy can also be protected, as encrypted data are stored in the cloud where the analysis is run. Currently, such data are stored in restricted access databases.

In addition to Eugene Koonin's review [14] on how the revolution in sequencing technologies has provided evolutionary insights into the tree of life and a review by Robert Plenge [15] that discusses the success of genome-wide association studies for determining the genetic basis of autoimmune diseases, overall, these reviews mark key genomic developments in the past ten years.

So here we are, ten years on from the announcement that the human genome had been sequenced. Rather

than considering this as an end point, however, we are now looking forward to a future with almost more data than we know what to do with and advances in our understanding of human biology and disease that will surely affect everyone. It seems that Clinton's pledge to drive forward research and develop translational therapies will be realized and that momentous occasion ten years ago was just the start of certainly the most exciting time in biological research.

Published: 5 May 2010

#### References

1. **Genome Biology** [<http://genomebiology.com/>]
2. Petsko GA: **And they said it wouldn't last...** *Genome Biology* 2010, 11:121.
3. **Genome Medicine** [<http://www.genomemedicine.com/>]
4. **Beyond the Genome** [<http://www.beyondthegenome2010.com>]
5. Perlea M, Salzberg SL: **Between a chicken and a grape: estimating the number of human genes.** *Genome Biology* 2010, 11:206.
6. Fischbach MA, Krogan NJ: **The next frontier of systems biology: higher order interactions.** *Genome Biology* 2010, 11:208.
7. **The NIH Common Fund** [<http://nihroadmap.nih.gov/hmp/>]
8. **MetaHIT Website** [<http://www.metahit.eu/index.php?id=410>]
9. Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE, Gordon JL, Knight R: **Study of the human microbiome by direct sequencing readily reveals community differences.** *Genome Biology* 2010, 11:210.
10. **The Cancer Genome Atlas** [<http://cancergenome.nih.gov/>]
11. **International Cancer Genome Consortium** [<http://www.icgc.org/>]
12. Mardis ER: **Cancer genomics identifies determinants of tumor biology.** *Genome Biology* 2010, 11:211.
13. Stein LD: **The case for cloud computing in genome informatics.** *Genome Biology* 2010, 11:207.
14. Koonin EV: **The origin and early evolution of eukaryotes in the light of comparative genomics.** *Genome Biology* 2010, 11:209.
15. Plenge R: **GWAS and the age of the human model organism for use in autoimmune genetic research.** *Genome Biology* 2010, 11:212.

doi:10.1186/gb-2010-11-5-120

**Cite this article as:** Garvey C: **A decade and genome of change.** *Genome Biology* 2010, 11:120.