doi:10.1038/nature10242

An integrated semiconductor device enabling non-optical genome sequencing

Jonathan M. Rothberg[1], Wolfgang Hinz[1], Todd M. Rearick[1], Jonathan Schultz[1], William Mileski[1], Mel Davey[1], John H. Leamon[1], Kim Johnson[1], Mark J. Milgrew[1], Matthew Edwards[1], Jeremy Hoon[1], Jan F. Simons[1], David Marran[1], Jason W. Myers[1], John F. Davidson[1], Annika Branting[1], John R. Nobile[1], Bernard P. Puc[1], David Light[1], Travis A. Clark[1], Martin Huber[1], Jeffrey T. Branciforte[1], Isaac B. Stoner[1], Simon E. Cawley[1], Michael Lyons[1], Yutao Fu[1], Nils Homer[1], Marina Sedova[1], Xin Miao[1], Brian Reed[1], Jeffrey Sabina[1], Erika Feierstein[1], Michelle Schorn[1], Mohammad Alanjary[1], Eileen Dimalanta[1], Devin Dressman[1], Rachel Kasinskas[1], Tanya Sokolsky[1], Jacqueline A. Fidanza[1], Eugeni Namsaraev[1], Kevin J. McKernan[1], Alan Williams[1], G. Thomas Roth[1] & James Bustillo[1]

The seminal importance of DNA sequencing to the life sciences, biotechnology and medicine has driven the search for more scalable and lower-cost solutions. Here we describe a DNA sequencing technology in which scalable, low-cost semiconductor manufacturing techniques are used to make an integrated circuit able to directly perform non-optical DNA sequencing of genomes. Sequence data are obtained by directly sensing the ions produced by template-directed DNA polymerase synthesis using all-natural nucleotides on this massively parallel semiconductor-sensing device or ion chip. The ion chip contains ion-sensitive, field-effect transistor-based sensors in perfect register with 1.2 million wells, which provide confinement and allow parallel, simultaneous detection of independent sequencing reactions. Use of the most widely used technology for constructing integrated circuits, the complementary metal-oxide semiconductor (CMOS) process, allows for low-cost, large-scale production and scaling of the device to higher densities and larger array sizes. We show the performance of the system by sequencing three bacterial genomes, its robustness and scalability by producing ion chips with up to 10 times as many sensors and sequencing a human genome.

DNA sequencing and, more recently, massively parallel DNA sequencing[1–4] has had a profound impact on research and medicine. The reductions in cost and time for generating DNA sequence have resulted in a range of new sequencing applications in cancer[5,6], human genetics[7], infectious diseases[8] and the study of personal genomes[9–11], as well as in fields as diverse as ecology[12,13] and the study of ancient DNA[14,15]. Although *de novo* sequencing costs have dropped substantially, there is a desire to continue to drop the cost of sequencing at an exponential rate consistent with the semiconductor industry's Moore's Law[16] as well as to provide lower cost, faster and more portable devices. This has been operationalized by the desire to reach the $1,000 genome[17].

To date, DNA sequencing has been limited by its requirement for imaging technology, electromagnetic intermediates (either X-rays[18], or light[19]) and specialized nucleotides or other reagents[20]. To overcome these limitations and further democratize the practice of sequencing, a paradigm shift based on non-optical sequencing on newly developed integrated circuits was pursued. Owing to its scalability and its low power requirement, CMOS processes are dominant in modern integrated circuit manufacturing[21]. The ubiquitous nature of computers, digital cameras and mobile phones has been made possible by the low-cost production of integrated circuits in CMOS.

Leveraging advances in the imaging field—which has produced large, fast arrays for photonic imaging[22]—we sought a suitable electronic sensor for the construction of an integrated circuit to detect the hydrogen ions that would be released by DNA polymerase[23] during sequencing by synthesis, as opposed to a sensor designed for the detection of photons. Although a variety of electrochemical detection methods have been studied[24,25], the ion-sensitive field-effect transistor (ISFET)[26,27] was most applicable to our chemistry and scaling requirements because of its sensitivity to hydrogen ions, and its compatibility with CMOS processes[28–31]. Previous attempts to detect both single-nucleotide polymorphisms (SNPs)[32] and DNA synthesis[33] as well as sequence DNA electronically[34] have been made. However, none of them produced *de novo* DNA sequence, addressed the issue of delivering template DNA to the sensors, or scaled to large arrays. In addition, previous efforts in ISFETs were limited in the number of sensors per array, the yield of working independent sensors and readout speed[35,36], and encountered difficulty in exposing the sensors to fluids while protecting the electronics[37].

Here, we overcome previous limitations with electronic detection and enable the production of chips with a large number of fast, uniform, working sensors. Our focus has been on the development of these ion chips, as well as the biochemical methods, supporting instrumentation and software needed to enable *de novo* DNA sequencing for applications requiring millions to billions of bases (Supplementary Fig. 1). A typical 2-h run using an ion chip with 1.2 M sensors generates approximately 25 million bases. The performance of the ion chips and overall sequencing platform is demonstrated through whole-genome sequencing of three bacterial genomes. The scalability of our chip architecture is demonstrated by producing chips with up to 10 times the number of sensors and producing a low-coverage sequence of the genome of Gordon Moore, author of Moore's law[16].

A CMOS integrated circuit for sequencing

We have developed a simple, scalable ISFET sensor architecture using electronic addressing common in modern CMOS imagers (Supplementary Fig. 2). Our integrated circuit consists of a large array of sensor elements, each with a single floating gate connected to an underlying ISFET (Fig. 1a). For sequence confinement we rely on a

[1]Ion Torrent by Life Technologies, Suite 100, 246 Goose Lane, Guilford, Connecticut 06437, USA.

QIAGEN Exhibit 1023
Page 1

©2011 Macmillan Publishers Limited. All rights reserved

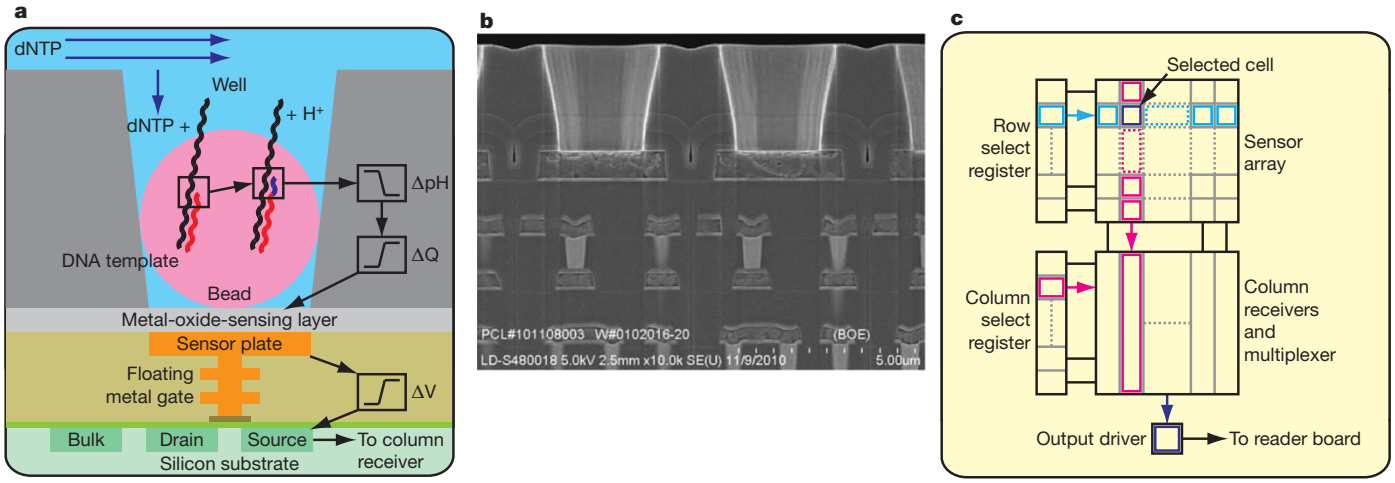

Figure 1 | Sensor, well and chip architecture. **a**, A simplified drawing of a well, a bead containing DNA template, and the underlying sensor and electronics. Protons (H^+) are released when nucleotides (dNTP) are incorporated on the growing DNA strands, changing the pH of the well (ΔpH). This induces a change in surface potential of the metal-oxide-sensing layer, and a change in potential (ΔV) of the source terminal of the underlying field-effect transistor. **b**, Electron micrograph showing alignment of the wells over the ISFET metal sensor plate and the underlying electronic layers. **c**, Sensors are arranged in a two-dimensional array. A row select register enables one row of sensors at a time, causing each sensor to drive its source voltage onto a column. A column select register selects one of the columns for output to external electronics.

3.5-μm-diameter well formed by adding a 3-μm-thick dielectric layer over the electronics and etching to the sensor plate (Fig. 1b). A tantalum oxide layer provides for proton sensitivity (58 mv pH^{-1}; ref. 38). High-speed addressing and readout are accomplished by the semiconductor electronics integrated with the sensor array (Fig. 1c). The sensor and underlying electronics provide a direct transduction from the incorporation event to an electronic signal. Unlike light-based sequencing technology, we do not use the elements of the array to collect photons and form a larger image to detect the incorporation of a base; instead we use each sensor to independently and directly monitor the hydrogen ions released during nucleotide incorporation.

Ion chips are manufactured on wafers (Fig. 2a), cut into individual die (Fig. 2b) and robotically packaged with a disposable polycarbonate flow cell that isolates the fluids to regions above the sensor array and away from the supporting electronics to provide convenient sample loading as well as electrical and fluidic interfaces to the sequencing instrument (Fig. 2c). Chips were designed and fabricated with 1.5 M, 7.2 M and 13 M ISFETs (Supplementary Fig. 3). On the basis of the placement of the flow cell on the sensor array, 1.2 M, 6.1 M and 11 M wells and sensors are exposed to fluids, with 99.9% of the sensors sensitive to pH and usable for DNA sequencing (Supplementary Fig. 4). Increasing the numbers of sensors per chip was first achieved by increasing the die area, from 10.6 mm × 10.9 mm to 17.5 mm × 17.5 mm, and then by increasing the density of the sensors by reducing the number of transistors per sensor from three to two. Chip density is limited by the selection of the CMOS node and the number of transistors per sensing element. Using a 0.35 μm CMOS node the minimum spacing for a three-transistor sensor is 5.1 μm and for a two-transistor sensor it is 3.8 μm (Supplementary Fig. 5). To understand further the limits on density, we show that 1.3 μm wells are readily manufactured, can be aligned to sensors, enable the generation of high-quality sequence (Supplementary Fig. 6) and can, using a 110 nm node, be fabricated with a spacing as small as 1.68 μm (Supplementary Fig. 7).

Sequencing on a semiconductor device

The all-electronic detection system used by the ion chip simplifies and greatly reduces the cost of the sequencing instrument (Supplementary

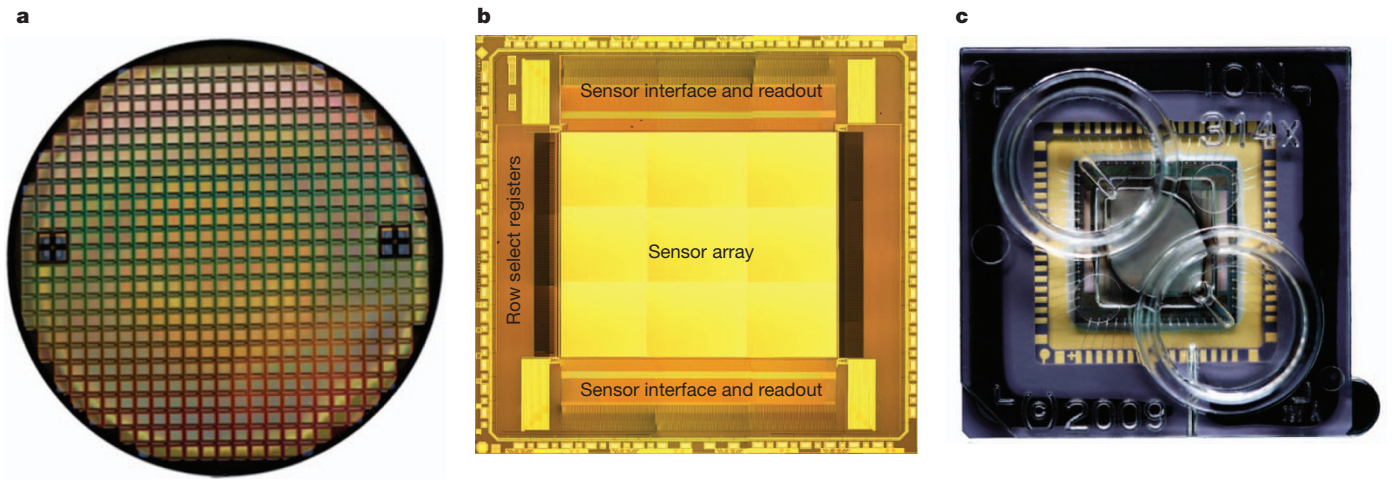

Figure 2 | Wafer, die and chip packaging. **a**, Fabricated CMOS 8′′ wafer containing approximately 200 individual functional ion sensor die. **b**, Unpackaged die, after automated dicing of wafer, with functional regions indicated. **c**, Die in ceramic package wire bonded for electrical connection, shown with moulded fluidic lid to allow addition of sequencing reagents.

QIAGEN Exhibit 1023
Page 2

©2011 Macmillan Publishers Limited. All rights reserved

Fig. 8). The instrument has no optical components, and is comprised primarily of an electronic reader board to interface with the chip, a microprocessor for signal processing, and a fluidics system to control the flow of reagents over the chip (Supplementary Fig. 9).

Genomic DNA is prepared for sequencing as described in Supplementary Methods. Briefly, DNA is fragmented, ligated to adapters, and adaptor-ligated libraries are clonally amplified onto beads. Template-bearing beads are enriched through a magnetic-bead-based process. Sequencing primers and DNA polymerase are then bound to the templates and pipetted into the chip's loading port. Individual beads are loaded into individual sensor wells by spinning the chip in a desktop centrifuge. A 2 μm acrylamide bead was chosen to deliver sufficient copies of the template to the sensor well to achieve a high signal-to-noise ratio (SNR) (800 K copies, SNR, 10; Supplementary Methods and Supplementary Fig. 10), while well depth was selected to allow only a single bead to occupy a well.

In ion sequencing, all four nucleotides are provided in a stepwise fashion during an automated run (Supplementary Methods). When the nucleotide in the flow is complementary to the template base directly downstream of the sequencing primer, the nucleotide is incorporated into the nascent strand by the bound polymerase. This increases the length of the sequencing primer by one base (or more, if a homopolymer stretch is directly downstream of the primer) and results in the hydrolysis of the incoming nucleotide triphosphate, which causes the net liberation of a single proton for each nucleotide incorporated during that flow. The release of the proton produces a shift in the pH of the surrounding solution proportional to the number of nucleotides incorporated in the flow (0.02 pH units per single base incorporation). This is detected by the sensor on the bottom of each well, converted to a voltage and digitized by off-chip electronics (Fig. 3). The signal generation and detection occurs over 4 s (Fig. 3b). After the flow of each nucleotide, a wash is used to ensure nucleotides do not remain in the well. The small size of the wells allows diffusion into and out of the well on the order of a one-tenth of a second and eliminates the need for enzymatic removal of reagents[1].

Signal processing and base calling

To change raw voltages into base calls, signal-processing software converts the raw data into measurements of incorporation in each well for each successive nucleotide flow using a physical model. Sampling the signal at high frequency relative to the time of the incorporation signal allows signal averaging to improve the SNR. The physical model takes into consideration diffusion rates, buffering effects and polymerase rates (Supplementary Fig. 11). The model is applied and fit to the raw trace from each well and the incorporation signals are extracted. A base caller corrects the signals for phase and signal loss, normalizes to the key, and generates corrected base calls for each flow in each well to produce the sequencing reads (Fig. 3c and Supplementary Fig. 12).

Next, each read is sequentially passed through two signal-based filters to exclude low-accuracy reads. The first filter measures the fraction of flows in which an incorporation event was measured. When this value is unusually large (greater than 60% of the first 60 flows) the read is not clonal. The second filter measures the extent to which the observed signal values match those predicted by the phasing model. When there is poor agreement (median absolute difference more than 0.06 over the first 60 flows) between the two, it corresponds to higher error rates. Lastly, per-base quality values are predicted using an adaptation of the Phred method[39] that quantifies the concordance between the phasing model predictions and the observed signal. These *ab initio* scores track closely with post-alignment derived quality scores, and are used to trim back low-quality sequence from the 3′ end of a read (Supplementary Fig. 13).

Sequencing bacterial genomes

Bacterial genome sequencing and signal processing was performed as described earlier. We succeeded in sequencing all three genomes fivefold to tenfold in individual runs using the small ion chip, covering 96.80% to 99.99% of each genome, with genome-wide consensus accuracies as high as 99.99% (Table 1 and Supplementary Fig. 14). *Escherichia coli* sequencing with three successively larger ion chips produced 46 to over 270 megabases of sequence (Table 1).

To characterize run quality, we aligned each read to the corresponding reference genome (Supplementary Fig. 15). The per-base accuracy was observed to be 99.569% ± 0.001% within the first 50 bases and 98.897% ± 0.001% within the first 100 bases (Supplementary Fig. 16a). This accuracy is similar at 50 bases and higher at 100 bases than light-based methods using modified nucleotides (1.1% versus 5% error[40]). The per-base accuracy in calling a homopolymer of length 5 is 97.328% ± 0.023% (Supplementary Fig. 16b) and higher than pyrosequencing-based sequencing methods[1,41]. For each genome, the observed distribution of per-base coverage matches closely with the theoretical Poisson distribution reflecting the uniform nature of the coverage (Supplementary Fig. 17). The distribution of coverage was also relatively unbiased across GC content (Supplementary Fig. 18).

Ion sequencing technology has allowed the routine acquisition of 100-base read lengths, and perfect read lengths exceeding 200 bases (Supplementary Fig. 19). At present, 20–40% of the sensors in a given

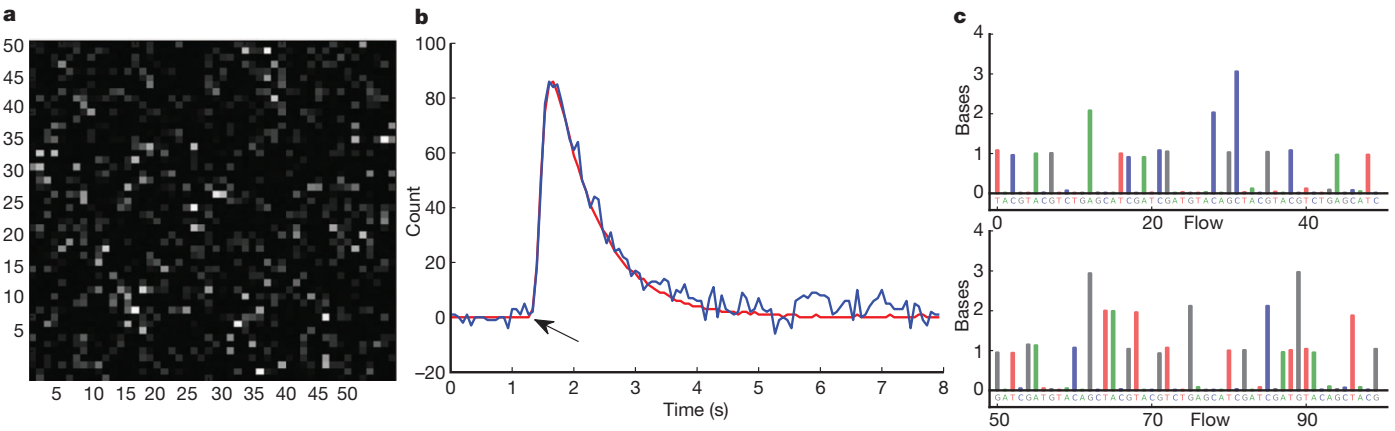

Figure 3 | Data collection and base calling. **a**, A 50 × 50 region of the ion chip. The brightness represents the intensity of the incorporation reaction in individual sensor wells. **b**, 1-nucleotide incorporation signal from an individual sensor well; the arrow indicates start of incorporation event, with the physical model (red line) and background corrected data (blue line) shown. **c**, The first 100 flows from one well. Each coloured bar indicates the corresponding number of bases incorporated during that nucleotide flow.

QIAGEN Exhibit 1023
Page 3

©2011 Macmillan Publishers Limited. All rights reserved

Table 1 | *Vibrio fisheri*, *E. coli*, *Rhodopseudomonas palustris* and *Homo sapiens*

	V. fisheri	*R. palustris*	*E. coli*	*E. coli*	*E. coli*	*H. sapiens*
GC content	38%	65%	51%	51%	51%	41%
Genome size	4.2 Mb	5.5 Mb	4.7 Mb	4.7 Mb	4.7 Mb	2.9 Gb
Number of runs x ion chip size	1 × 1.2 M	1 × 1.2 M	1 × 1.2 M	1 × 6.1 M	1 × 11 M	1,601 × 1.2 M 267 × 6.1 M 28 × 11.1 M
Fold coverage	6.2-fold	6.9-fold	11.3-fold	36.2-fold	58.4-fold	10.6-fold
Coverage	96.80%	99.64%	99.99%	100.00%	100.00%	99.21%
Reads ≥21 bases	261,313	444,750	507,198	1,852,931	2,594,031	366,623,578
Reads ≥50 bases	233,049	399,360	487,420	1,698,852	2,343,880	306,042,650
Reads ≥100 bases	156,391	160,726	400,743	1,012,918	1,779,237	139,624,090
Mapped bases	26.0 Mb	37.8 Mb	47.6 Mb	169.6 Mb	273.9 Mb	30.2 Gb

Coverage shows percentage of genome covered based on one or more reads mapping to each base of the reference genome. Reads align with 98% or greater accuracy.

run yield mappable reads. The gap between the number of sensors on a chip and the number yielding sequence is primarily the result of incomplete loading of the chip, poor amplification of a fragment onto the bead, and lack of clonality of the template. With continued improvements in loading and template preparation, along with improvements in signal processing and base calling, it is expected that the percentage of sensors yielding reads, the average read length and read accuracy will all improve significantly, as it has for other sequencing technologies[1–4,9–11].

'Post-light' sequencing of G. Moore

To illustrate the scalability of semiconductor sequencing we produced whole-genome sequence data from an individual, G. Moore[42] (Fig. 4). Written consent was provided by G. Moore to sequence and publish his genome and resulting findings. Reads from his genome were deposited in the Sequence Read Archive (SRA) under accession number ERP000682. The mean coverage of the G. Moore genome was 10.6-fold (Table 1). The degree to which the observed distribution of reads conforms to a Poisson distribution is indicative of a general lack of bias in coverage depth (Fig. 4b).

We found 2,598,983 SNPs in the G. Moore genome, of which 3.08% were found to be novel, consistent with previous reports[4,9,11] (Supplementary Methods). To confirm the accuracy of our analysis, we also sequenced the G. Moore genome using ABI SOLiD Sequencing[43] to 15-fold coverage and validated 99.95% of the heterozygous and 99.97% of the homozygous genotypes (Supplementary Tables 1 and 2).

We used the Online Mendelian Inheritance in Man database[44] and the 23andMe functional SNP collection (https://www.23andme.com) to identify a subset of validated SNPs known to be involved in human disease and interesting phenotypes (Supplementary Table 3). We also examined the G. Moore sequence for the 7,693 deletions and inversions discovered by the 1000 Genomes Consortium and computationally found 3,413 of them in the G. Moore genome at a 99.94% positive predictive value (Supplementary Methods, Supplementary Table 4 and Supplementary Fig. 20). To determine G. Moore's maternal ancestry, reads were also mapped to human mitochondrial DNA[45] for a mean coverage of 732-fold. G. Moore's mitochondria belong to haplogroup H, the most common in Europe[46].

Discussion

We have demonstrated the ability to produce and use a disposable integrated circuit fabricated in standard CMOS foundries to perform, for the first time, 'post-light' genome sequencing of bacterial and human genomes. With fifty billion dollars spent per year on CMOS semiconductor fabrication and packaging technologies, our goal was to leverage that investment to make a highly scalable sequencing technology. Using the G. Moore genome we demonstrated the feasibility of sequencing a human genome. The G. Moore genome sequence required on the order of a thousand individual ion chips comprising about one billion sensors. By demonstrating the ability to make larger and denser arrays, use fewer transistors per sensor, and sequence from wells as small as 1.3 μm, our work suggests that readily available CMOS nodes should enable the production of one-billion-sensor ion chips and low-cost routine human genome sequencing.

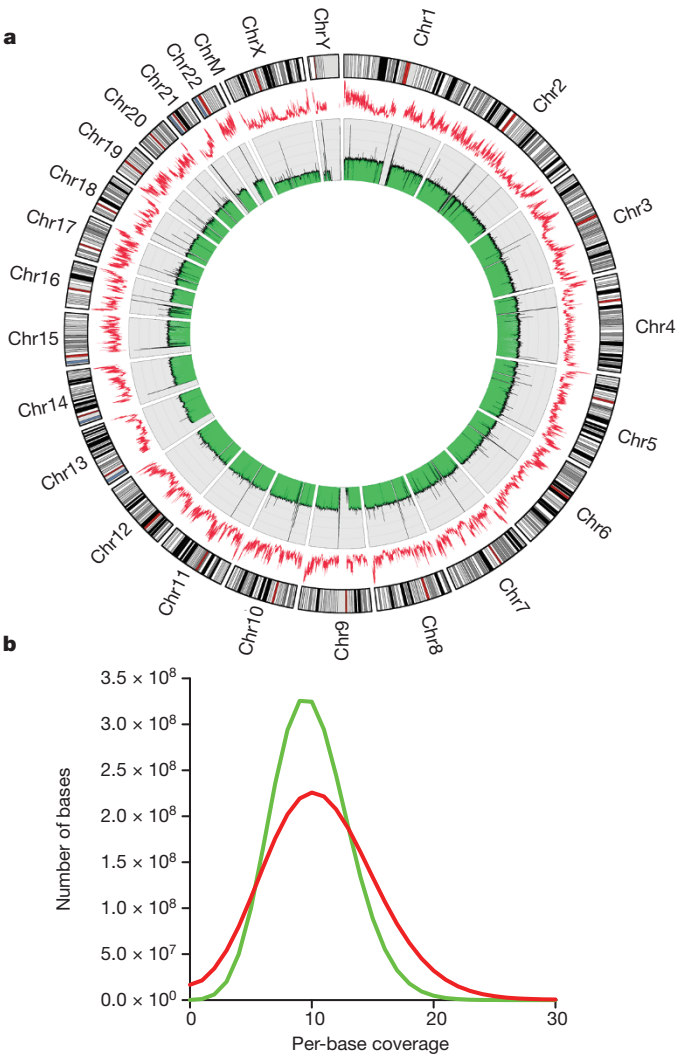

Figure 4 | G. Moore genome. **a**, Circular genome plot. The average sequencing coverage (green) and average GC content (red) within 100-kb intervals is shown. **b**, Distribution of the observed per-base coverage depth along the genome (red) compared with the distribution expected from random coverage (green).

QIAGEN Exhibit 1023
Page 4

©2011 Macmillan Publishers Limited. All rights reserved

Received 8 March; accepted 26 May 2011.

1. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
2. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309,** 1728–1732 (2005).
3. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
4. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327,** 78–81 (2010).
5. Thomas, R. K. *et al.* Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nature Med.* **12,** 852–855 (2006).
6. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456,** 66–72 (2008).
7. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genet.* **42,** 30–35 (2010).
8. Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307,** 223–227 (2005).
9. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452,** 872–876 (2008).
10. Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. *N. Engl. J. Med.* **362,** 1181–1191 (2010).
11. Schuster, S. C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463,** 943–947 (2010).
12. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
13. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl Acad. Sci. USA* **103,** 12115–12120 (2006).
14. Noonan, J. P. *et al.* Genomic sequencing of Pleistocene cave bears. *Science* **309,** 597–599 (2005).
15. Green, R. E. *et al.* Analysis of one million base pairs of Neanderthal DNA. *Nature* **444,** 330–336 (2006).
16. Moore, G. E. Cramming more components onto integrated circuits. *Electronics* **38,** 114–117 (1965).
17. Davies, K. *The $1,000 Genome* (Free Press, 2010).
18. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74,** 5463–5467 (1977).
19. Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679 (1986).
20. Metzker, M. L. Sequencing technologies—the next generation. *Nature Rev. Genet.* **11,** 31–46 (2010).
21. Wanlass, F. M. Low stand-by power complementary field effect circuitry. U.S. patent 3,356,858: (1967).
22. Theuwissen, A. J. P. CMOS image sensors: state-of-the-art. *Solid-State Electron.* **52,** 1401–1406 (2008).
23. Sakurai, T. & Husimi, Y. Real-time monitoring of DNA polymerase reactions by a micro ISFET pH sensor. *Anal. Chem.* **64,** 1996–1997 (1992).
24. Fritz, J., Cooper, E. B., Gaudet, S., Sorger, P. K. & Manalis, S. R. Electronic detection of DNA by its intrinsic molecular charge. *Proc. Natl Acad. Sci. USA* **99,** 14142–14146 (2002).
25. Drummond, T. G., Hill, M. G. & Barton, J. K. Electrochemical DNA sensors. *Nature Biotechnol.* **21,** 1192–1199 (2003).
26. Bergveld, P. Development of an ion-sensitive solid-state device for neurophysiological measurements. *IEEE Trans. Biomed. Eng.* **17,** 70–71 (1970).
27. Bergveld, P. Thirty years of ISFETOLOGY. What happened in the past 30 years and what may happen in the next 30 years. *Sens. Actuators B Chem.* **88,** 1–20 (2003).
28. Yeow, T., Haskard, M., Mulcahy, D., Seo, H. & Kwon, D. A very large integrated pH-ISFET sensor array chip compatible with standard CMOS processes. *Sens. Actuators B Chem.* **44,** 434–440 (1997).
29. Bausells, J., Carrabina, J., Errachid, A. & Merlos, A. Ion-sensitive field-effect transistors fabricated in a commercial CMOS technology. *Sens. Actuators B Chem.* **57,** 56–62 (1999).
30. Milgrew, M., Hammond, P. & Cumming, D. The development of scalable sensor arrays using standard CMOS technology. *Sens. Actuators B Chem.* **103,** 37–42 (2004).
31. Hizawa, T., Sawada, K., Takao, H. & Ishida, M. Fabrication of a two-dimensional pH image sensor using a charge transfer technique. *Sens. Actuators B Chem.* **117,** 509–515 (2006).
32. Purushothaman, S., Toumazou, C. & Ou, C. P. Protons and single nucleotide polymorphism detection: A simple use for the ion sensitive field effect transistor. *Sens. Actuators B Chem.* **114,** 964–968 (2006).
33. Pourmand, N. *et al.* Direct electrical detection of DNA synthesis. *Proc. Natl Acad. Sci. USA* **103,** 6466–6470 (2006).
34. Sakata, T. & Miyahara, Y. DNA sequencing based on intrinsic molecular charges. *Angew. Chem.* **118,** 2283–2286 (2006).
35. Milgrew, M. J., Riehle, M. O. & Cumming, D. R. S. A large transistor-based sensor array chip for direct extracellular imaging. *Sens. Actuators B Chem.* **111–112,** 347–353 (2005).
36. Milgrew, M. J. & Cumming, D. R. S. Matching the transconductance characteristics of CMOS ISFET arrays by removing trapped charge. *Electron Devices. IEEE Trans. Electron Devices* **55,** 1074–1079 (2008).
37. Hammond, P. & Cumming, D. Encapsulation of a liquid-sensing microchip using SU-8 photoresist. *Microelectron. Eng.* **73–74,** 893–897 (2004).
38. Mikolajick, T., Kühnhold, R. & Ryssel, H. The pH-sensing properties of tantalum pentoxide films fabricated by metal organic low pressure chemical vapor deposition. *Sens. Actuators B Chem.* **44,** 262–267 (1997).
39. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).
40. Claesson, M. J. *et al.* Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* **38,** e200 (2010).
41. Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L. & Mark Welch, D. Accuracy and quality of massively-parallel DNA pyrosequencing. *Genome Biol.* **8,** R143 (2007).
42. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).
43. McKernan, K. J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19,** 1527–1541 (2009).
44. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33,** D514–D517 (2005).
45. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genet.* **23,** 147 (1999).
46. Kloss Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32,** 25–32 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We want to thank G. Moore for his willingness to participate in this study. We thank G. Fergus, M. Jain, J. Kole, L. Stevens and the ION team for supporting our efforts, and H. Peckman, V. Tadigotla, D. Holloway and S. Mclaughlin for help on the variant analysis, and M. Ross of the Broad Institute for help on quality scores. This research was supported, in part, by a grant from the National Human Genome Research Institute (NHGRI), RFA-HG-08-008, Revolutionary Genome Sequencing Technologies—The $1000 Genome. Grant number: R01 HG005094.

Author Contributions J.M.R. conceived the technology, supervised the project and wrote the manuscript with input from co-authors. K.J., M.J.M. and J.B. designed chips. J.F.D., M.A., D.L., J.W.M., J.F.S., E.N., M.S., X.M., A.B., T.A.C., M.H., I.B.S., B.R., J.S., E.F., M.S., J.A.F., K.J.M. and J.H.L. developed methods. M.D., J.T.B., M.E., J.H., N.H., T.M.R., B.P.P., S.E.C., M.L., Y.F. and A.W. wrote software and analysed data. W.H., J.S., W.M., D.M., J.R.N. and G.T.R. designed the instrument. E.D., D.D., R.K. and T.S. sequenced the human sample.

Author Information Sequences for *Homo sapiens*, *Escherichia coli*, *Vibrio fisheri* and *Rhodopseuomanas palustris* were deposited in the Sequence Read Archive (SRA) under accession numbers ERP000682, ERP000541, ERP000542 and ERP000543 respectively. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike Licence, and is freely available to all readers at www.nature.com/nature. The authors declare competing financial interests: details accompany the full-text HTML version of this paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.R. (Jonathan.Rothberg@Lifetech.com).

QIAGEN Exhibit 1023
Page 5

©2011 Macmillan Publishers Limited. All rights reserved