WS AMATI GLOBAL INNOVATION FUND

Innovation Frontier

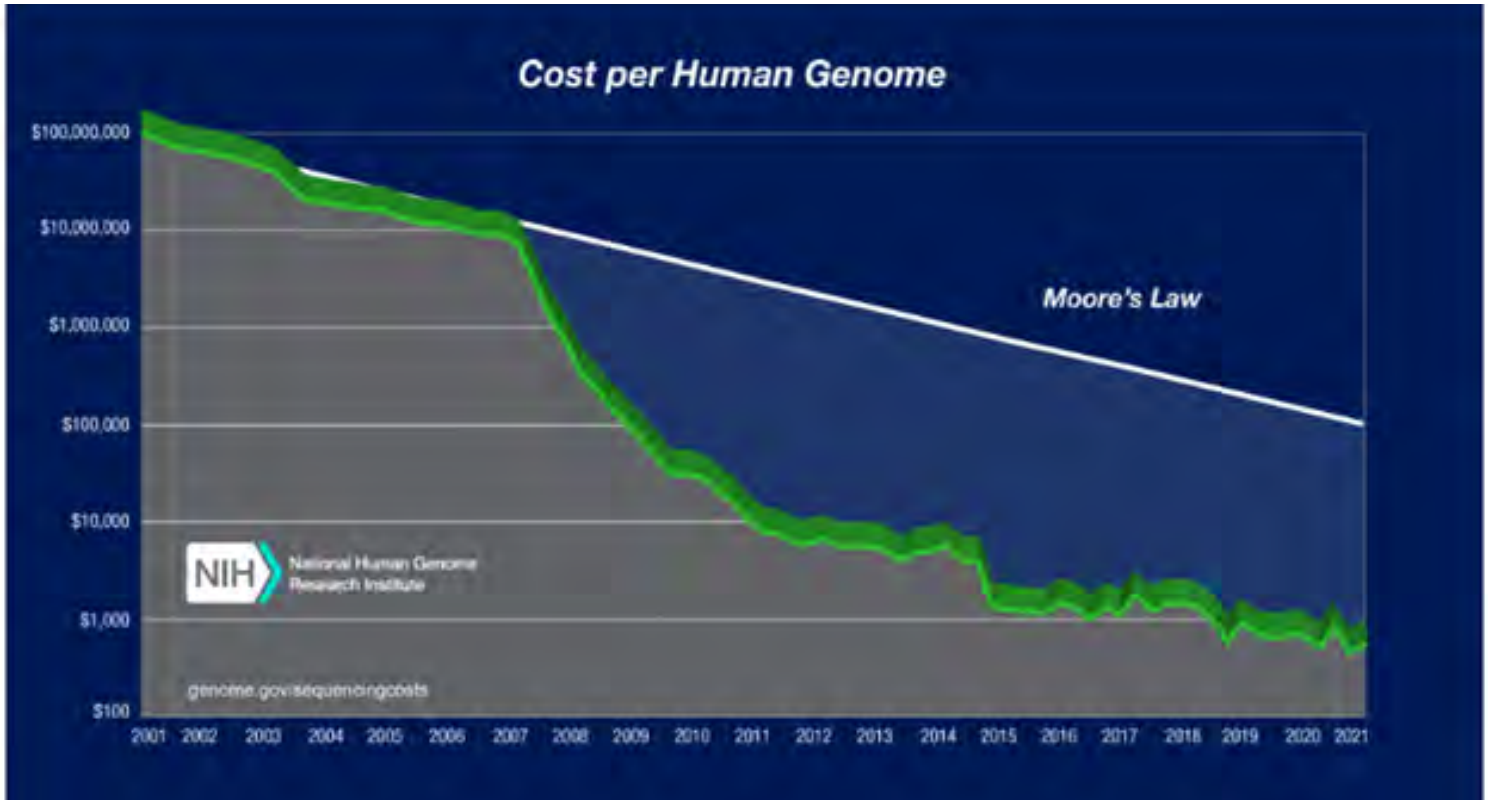
Genetic Sequencing

By Dr Gareth Blades, Analyst

Amati
Global Investors

The Human Genome Project (HGP) was conducted over 13 years with the first sequenced human genome published as a draft in 2000 and finalised in 2003. It is reported to have cost $3bn over the course of the project. Contrast that with today, where companies are touting human genome sequencing in less than 20 hours at a cost of $200 per genome. This is all the more impressive when you dig into the detail. Technically, the Human Genome Project sequenced only one of each base pair - half of the human genome - to produce a draft sequence in 10 years. Technology now used daily enables the sequencing of both base pairs, one from each strand of the double helix. This is referred to as whole genome sequencing (WGS). What's more, multiple (>100) whole genomes can be sequenced in parallel on the highest throughput machines.

Figure 1: The drop in sequencing costs from early 2008 is due to the introduction of second generation or "next generation" sequencing (NGS).



Source: The National Human Genome Research Institute

The developments demonstrated in Figure 1, and the subsequent drop with the current generation of machines, has caused a proliferation and expansion of sequencing into new domains.

There is a vision emerging in the industry of having sequencing at every point in the healthcare continuum; from family planning through to disease recurrence testing. The unifying element of these applications is the growth of clinical testing, the most valuable part of the market. For example, those considering starting a family may want to test themselves for rare disease variants they might pass on; whilst pregnant, this could be followed with non-invasive prenatal testing for chromosomal number disorders; disease screening using next generation sequencing (NGS) could be used in early and more accurate diagnosis; if diagnosed, a therapy that has been developed with the aid of NGS, could be selected for maximal efficacy based on your detailed genetic background; the effectiveness of the therapy could be confirmed with, you guessed it, NGS; and once in remission, recurrence monitoring can take place with NGS.

What is required for this to become a reality? I'd argue: cost, data, analysis and patient access.

Cost

Even cheaper sequencing is required. $200 is still prohibitively high for mass adoption if it is being truly democratised. Digging into this figure, Illumina which is the largest provider of NGS machines, for its flagship sequencer is quoting $200 per genome, including the cost of secondary analysis. The newly launched flagship series, the NovaSeqX/X+, are designed for the highest throughput labs, running 10s of 1000s of genomes per year. These machines include both computing hardware and data processing workflows and this integration saves a significant amount of time at the end of a sequencing run. The time saving is most significant for the highest throughput users with these highly efficient, and very expensive, machines simply not economical for lower volume users.

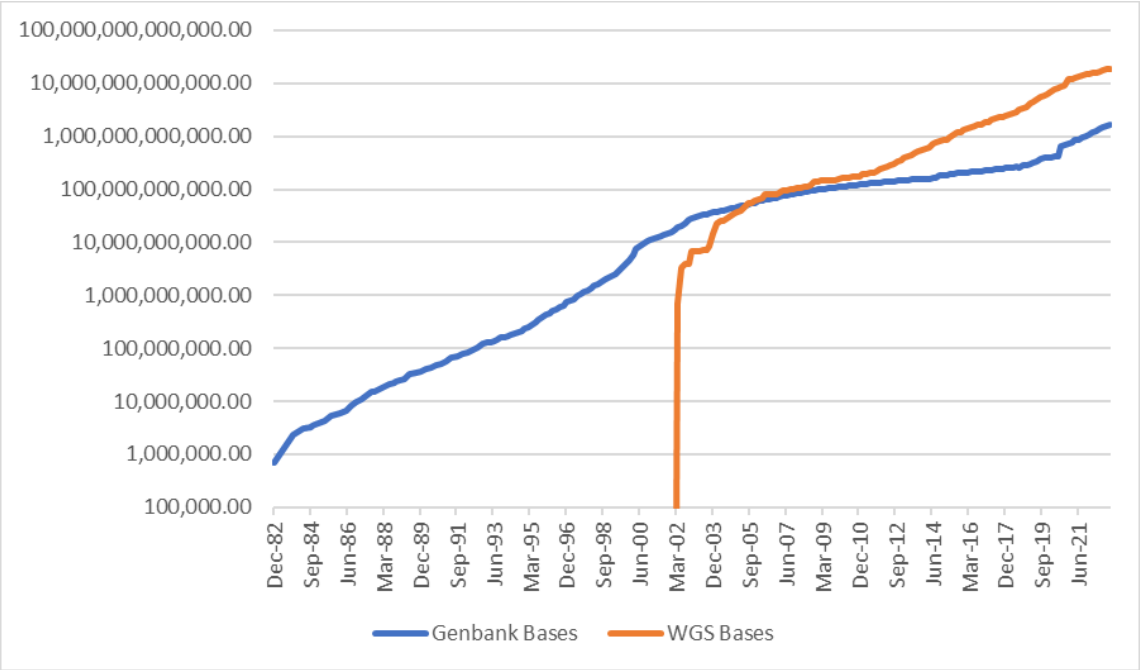
Not all sequencers include analysis in the headline price though. In older and lower throughput machines, data is transferred to cloud or locally before it can be manipulated and processed. This adds time and cost. The private company, Element Biosciences, offers the AVITI sequencer for mid-throughput users, with a cost per genome of $200, not including analysis. AVITI runs on a subscription programme, that becomes cheaper per run as use of the machine scales. While another private company, Ultima Genomics, is touting a $100 genome, although further details are scant.

While the cost per genome is an important figure, it is worth remembering that when genomic insights are productised into tests, they rely on sequencing small portions of the genome rather than the whole. They can also move from NGS machines to more commonplace and cheaper PCR platforms or other test modalities depending on what is being assayed. What is more, clinical tests tend to cost a premium above the cost per genome. For oncology tests, this can run into the low thousand dollars.

Data

Regardless of whether the lab is doing 1000 or 10s of 1000s of genome per year, the outcome is more genomic data. A single human genome sequence creates 200 gigabytes of data. It is estimated that between 2bn - 40bn gigabytes of genomic data is generated each year1. At the top end, this is equivalent to 200m individual human genomes, although this figure will include data from animals, plants, fungi, bacteria, viruses etc.

Figure 2: Cumulative number of sequenced bases and whole genome sequencing (WGS) bases uploaded to public database, GenBank. Reproduced from GenBank database



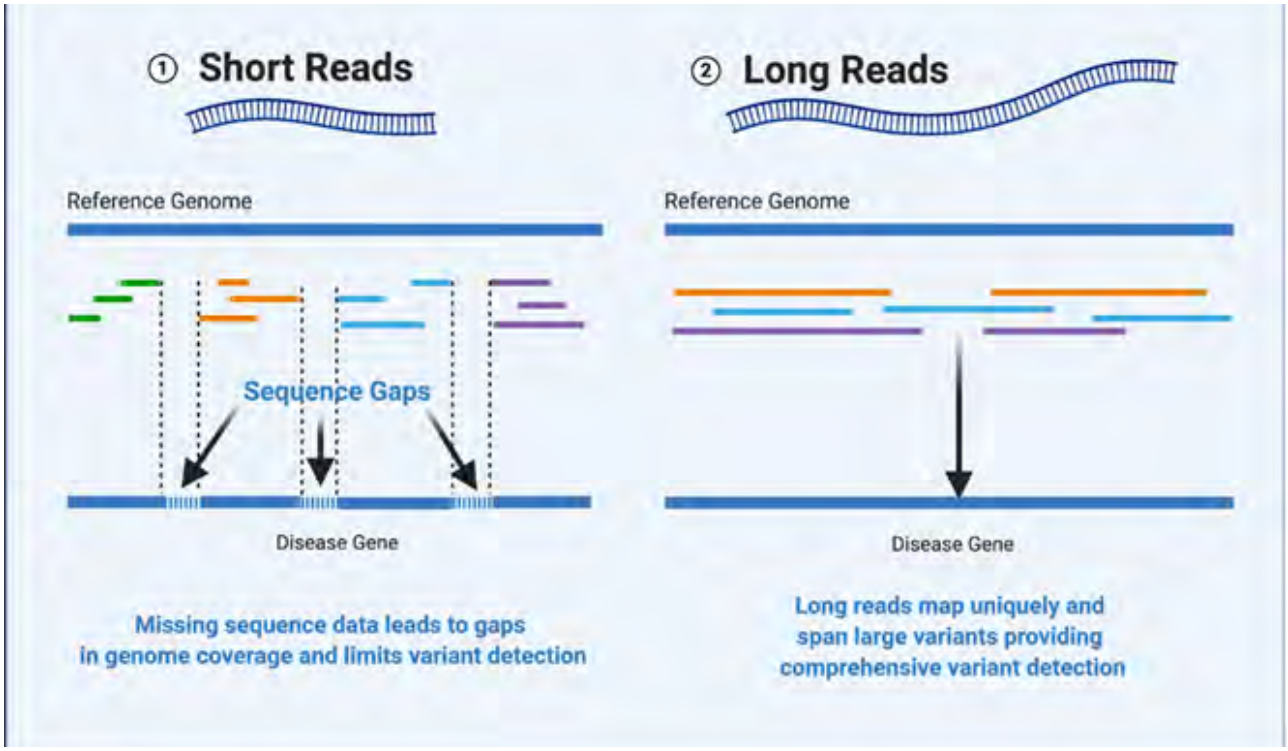
Source: National Library of Medicene

This amount of genomic data could be a nightmare scenario for some and leads to debates about data stewardship and ownership. However, it is what this data is composed of and from whom that is the problem for me. On this second point, the majority of sequencing projects to date have been done in affluent western countries with predominantly white participants. There is a paucity of data about any other ethnic group. This skews the data and any insights that can be drawn from it. For example, understanding disease onset risk for screening programmes would under-represent any section of the population not captured in the raw data set.

On the first point, we must dig into how sequencing reads are generated. There are broadly speaking, two competing technologies; short read and long read. Short read sequencing is just that, the genome is split into short sections (150-250 bases). These short sections are read multiple times and in overlapping increments. Software is used to align the overlapping parts of the sequences to build up longer and longer sections of the genome. Once this is complete a consensus sequence is the result. Long reads aim to read long lengths (10,000s bases) of the genome at once.

The long and short of it is, short read struggles to read certain areas of the genome. Some areas of the genome are "dark" because they have structural variations that mess up the certainty with which short read bases are called and subsequently aligned. Consensus sequences then don't make sense in these regions because the short reads can't be stitched together.

Figure 3: Schematic of short and long read technique accuracy for structural variant regions



Source: Hudson Alpha, Institute for Biotechnology

This is a two-fold problem, firstly it is suggested that regions with structural variations contain crucial genetic information such as disease-causing modifications. The second part of this problem is that there is so much short-read data that it is much easier to keep working with this technology because the existing and ever-expanding dataset allows more meaningful insights to be generated. This is why there are more clinical applications of short read. The paucity of long-read data makes clinical insights harder to generate and will thus miss important insights into the health of the population. This is likely to remain the case until long-read systems, such as from PacBio or Oxford Nanopore, reach the scale of install and acceptance as Illumina's. PacBio has an installed base of ~1000, while Illumina has an installed base of ~23,000 machines.

Analysis

Downstream, all this data needs processing to generate meaningful clinical insights. For example, to understand the importance of different genomic variants, what risk of disease onset might be or what potential drug development avenues could be followed. In addition, downstream of basic research when clinical decisions need to be made and smaller parts of the genome are examined, how can the result of these genomic panels be interpreted accurately? Ready to use workflows and services already exist to place the genomic data in its biomedical context. Qiagen, a holding in the fund, among other things offers a suite of services that put genomic data in the context of its biological and clinical niche. Downstream effects can be predicted as can new drug targets or candidate biomarkers.

The machine manufacturers, Illumina, PacBio, Oxford Nanopore, Element and Ultima, all use different chemistries that are particular to their machines. Qiagen's suite of analysis tools and AI are compatible with the output from all these machines. While it is true Illumina is integrating analysis, Qiagen's offering is the leading solution because of its breadth and clinical insight. In NGS/Bioinformatics, Qiagen is the leading provider by revenue, having twice the sales of the number two and four times the revenue of the number three provider. It is this area of innovation and growth that I think is under-appreciated in NGS. This is a view supported by the recent news that Qiagen is looking at options to increase investment in this business through the sale of a minority stake.

Value is being ascribed to the pioneers of sequencing because of the foundational tools they supply. However, without being able to understand how that genomic data fits with a complex biological hypothesis of disease or screening or diagnosis or response to therapy, there is little hope of enabling the scenario that sees sequencing or testing at every part of an individual's healthcare journey.

Patient Access

Once disease insights have been productised into kits and tests, they need to get to the patient. LabCorp, another fund holding, is an expert at test productization and particularly patient access, and patient access on a massive scale. A consequence of the pandemic has been a broadening of where testing takes place, whether that is at home, in diagnostic hubs, pharmacies or "in store" in the US. Moreover, patients have got used to being able to order tests online or through an app and take the test at their convenience. LabCorp has access to all of these either directly through their patient app and website or with partnerships with companies like Walgreens, a US pharmacy chain. More recently, LabCorp has completed acquisitions of hospital testing infrastructure both centrally and in the community. Consequently, LabCorp has successfully located itself everywhere testing can happen. On top of which, LabCorp's tests cover every healthcare decision from prenatal planning to disease recurrence. Perhaps the vision of sequencing across the healthcare continuum is already a reality?

Sales Team Contacts

Rachel Le Derf
Head of Sales & Marketing
rachel.lederf@amatiglobal.com
07979601223

Colin Thomson
Head of Intermediary Distribution
Northern England, Scotland & NI
colin.thomson@amatiglobal.com
07884026517

Jonathan Woolley
Sales Director
London, Midlands, SW England & Wales
jonathan.woolley@amatiglobal.com
07818203013

Thomas Whitfield
Sales Director
London & SE England
thomas.whitfield@amatiglobal.com
01315039120

Samantha Dalby
Sales and Insights Manager
samantha.dalby@amatiglobal.com
+44 (0) 131 503 9116

Olivia Pattison
Senior Sales Support Executive
olivia.pattison@amatiglobal.com
+44 (0) 131 503 9126

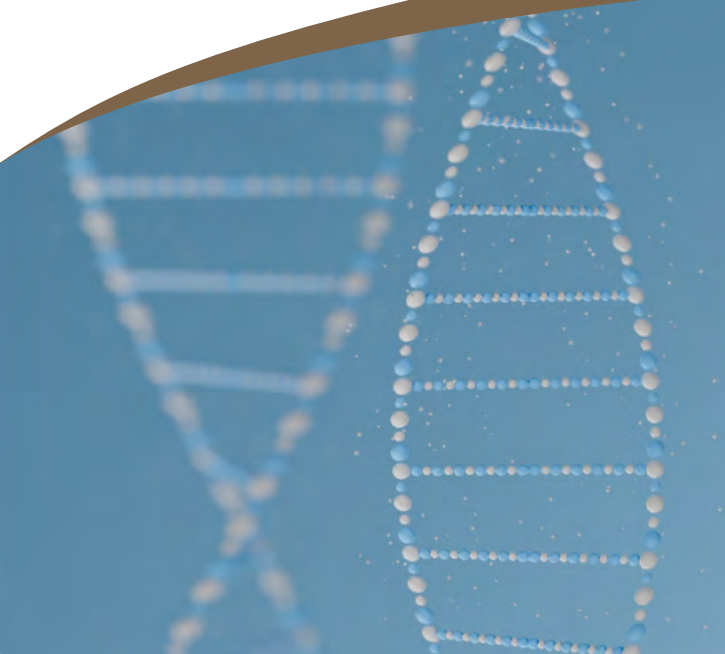
Milly Stevenson
Sales Support Executive
milly.stevenson@amatiglobal.com
+44 (0) 131 503 9125

Risk Warning

This article is a financial promotion issued by Amati Global Investors Limited, which is authorised and regulated by the Financial Conduct Authority. It is provided for informational purposes only and does not represent an offer or solicitation to buy or sell any securities; nor does it provide all the facts needed to make an informed decision about investing in the Fund. Prospective investors should always read the Key Investor Information Document and the Prospectus, which contain full details of the costs and charges applicable to the Fund as well as specific risk warnings. These documents are available at www.amatiglobal.com.

Past performance is not a reliable guide to future performance. The value of investments and the income from them may go down as well as up and is not guaranteed; investors may not get back the amount originally invested.

The return on investments in overseas markets may increase or decrease as a result of exchange rate movements. There may be occasions where there is an increased risk that a position in the Fund cannot be liquidated in a timely manner and at a reasonable price. In extreme circumstances this may affect the ability of the Fund to meet redemption requests upon demand.



Amati
Global Investors

Amati Global Investors Ltd

8 Coates Crescent, Edinburgh EH3 7AL

+44 (0)131 503 9115
info@amatiglobal.com
www.amatiglobal.com
Amati Global Investors is authorised and regulated by the Financial Conduct Authority.
Calls are recorded and monitored.