



**House of Lords Science & Technology Committee:
Sub-Committee II**

Inquiry into Genomic Medicine

**Submission from 454 Life Sciences, a Roche Company
10th November 2008**

EXECUTIVE SUMMARY

1. 454 Life Sciences, a Roche company, develops and commercialises the innovative Genome Sequencer™ system for ultra-high throughput DNA sequencing.
2. In response to the questions raised at the Sub-Committee's November 5th hearing concerning next-generation sequencing, 454 Life Sciences would like to highlight the points below
 - a) High throughput sequencing technologies have enabled an explosion of molecular biology research which is building a foundation for genomic medicine.
 - b) Government investment should be focused on the infrastructure required to translate the scientific understanding enabled by high-throughput sequencing into true clinical benefits. This should primarily be in the form of IT infrastructure and ensuring that there are enough trained individuals to manage the systems put in place, and providing the support for the development of nationwide databases.
 - c) It is important that clear guidelines are developed on the use of diagnostics, to ensure that new technology is assessed appropriately and introduced quickly and efficiently into the healthcare system where it can deliver real benefits for patients
 - d) The investment in projects such as the 1,000 Genome Project will benefit society - clinical benefits from high-throughput sequencing are imminent and likely to be realised well before the \$1,000 Genome

INTRODUCTION

About 454 Life Sciences

3. Based in Basel, Switzerland, Roche has grown over the past 100 years from a small pharmaceutical laboratory into one of the world's leading research-focused companies in the healthcare sector. Today, Roche employs around 75,000 people and operates in over 150 countries. The business is focused on two areas: diagnostics and pharmaceuticals.

4. Roche Applied Sciences is a global business area within the Diagnostics Division, employing approximately 1,800 worldwide. RAS's headquarters are in Penzberg, Germany; the UK operation is in Burgess Hill, West Sussex.

5. 454 Life Sciences, a centre of excellence of Roche Applied Sciences, develops and commercialises the innovative Genome Sequencer™ system for ultra-high-throughput DNA sequencing. Specific applications include *de novo* sequencing and re-sequencing of genomes, metagenomics, RNA analysis, and targeted sequencing of DNA regions of interest. 454



Sequencing technology has enabled over 250 peer-reviewed studies in diverse research fields including, cancer and infectious disease research, drug discovery and many more.

6 454 Life Sciences welcomes the opportunity to submit written evidence to the Sub-Committee. While the Roche Group and RAS have previously submitted memoranda, this submission specifically addresses the questions concerning next-generation sequencing which were raised in the Sub-Committee's November 5th hearing.

Advances in DNA Sequencing

7. For 30 years, the DNA sequencing method developed by Fred Sanger, dideoxynucleotide chain termination DNA sequencing, was the dominant method for DNA sequencing. Incorporating sequencing into genomic projects was unaffordable and often technically unfeasible, confining research to large scale genome centres. The launch of the Genome Sequencer 20 System by 454 Life Sciences System in 2005 put sequencing into the hands of a much broader research community by dramatically reducing the cost and time to complete genomic projects. Key advances underlying 454 Sequencing technology are massively parallel sequencing of DNA molecules in picoliter wells and a simple, unbiased method for sample preparation. Since 2005, 454 Life Sciences has released three generations of its technology which have continued to improve the cost and speed of sequencing while maintaining the industry-standard in quality.

8. The result of new, affordable sequencing technologies has been a grassroots groundswell of researchers that are suddenly finding high-throughput sequencing within their reach as a research tool. Researchers have published over 250 high quality publications in major journals including Nature and Science using the Genome Sequencer system. Many of these publications have had direct applicability to human health research and development. These studies cover many areas that will continue to be important in the characterisation and management of disease. Some major research areas covered by these publications include:

a) **Human Genome Characterisation:** The genome of Dr. James Watson was sequenced using long-read high-throughput sequencing, the first individual genome to be sequenced using a next generation technology. Several other papers have been published that characterise structural variations and chromosome conformations. The detailed characterisation of the variation amongst human genomes will make it possible to rapidly identify regions that are divergent in the germline DNA of patients with disease.

b) **Transcriptome Characterisation:** The transcriptome is the gene-containing part of the genome that is transcribed into RNA and then translated into proteins. Several publications have used long-read high-throughput sequencing to comprehensively and accurately sequence the transcriptomes of cancer cells in order to identify the genetic changes associated with oncogenesis. Studies such as these allow researchers to pinpoint genes that are defective and how they are altered in particular disease states.

c) **Target Genomic Region Characterisation:** Long-read high-throughput sequencing has been used to further characterise those parts of genomes that had previously been associated with disease. The research communities around the world have spent significant resources characterising the genetic basis of disease. This has been done with so-called whole genome association that results in pinpointing a region as having an association with a sub-population of the disease studied. However, these regions are typically large and don't help pharmaceutical researchers. Application of long-read high-throughput sequencing can allow researchers to quickly determine where in a disease associated region the important mutations reside, the so-called "causative alleles", by accurately sequencing the region in numerous patients.



d) **Targeted Gene Characterisation:** Long-read high-throughput sequencing has also been used to deeply characterise a variety of genes involved in disease process as well as those that are targets for pharmaceutical compounds. Characterisation of the P53, EGFR and the HIV Reverse Transcriptase (RT) loci are prime examples of the application of ultradeep long-read high-throughput sequencing. This deep sequencing of important genetic loci allows for the discovery or determination of rare mutations that may result in adverse events during drug treatment. Additionally studies such as these may suggest a better course of treatment for an individual patient thus reducing the overall cost of treatment or point to new targets for drug development.

e) **Virus Characterisation and Identification:** The study of infectious agents such as HIV has benefited greatly from the use of long-read high-throughput sequencing. Studies have been published that characterise HIV's integration in the human genome as part of its life cycle as well as mutational profiling studies of HIV before and during treatment with antiretroviral drugs.

In addition to the characterisation of known disease-causing virus, the Genome Sequencer System was used to identify an infectious virus that emerged in transplant recipient patients. Furthermore, the technology was used to uncover a novel viral fusion protein in a Merkel cell carcinoma in a study that suggests a path to treatment of this cancer. Other groundbreaking studies suggest that gut flora may play a larger role in obesity than previously thought.

f) **Other Infectious Agents:** Long-read high-throughput sequencing has been used to characterise the genomes of dozens of bacterial species many of which are important human pathogens including Mycobacterium tuberculosis (TB), Staphylococcus aureus (MRSA), Helicobacter, Streptococcus, Salmonella, Campylobacter, Bacillus anthracis and E. coli. The analysis of pathogen genomes allows for an understanding of the epidemiology of the infection, characterisation of drug resistance, and facilitates the development of new classes of antibiotic drugs.

9. There are now several next-generation high-throughput sequencing technologies commercially available. Each offers a unique combination of throughput and data quality. For example, 454 Sequencing generates relatively long read lengths (500 bases) with very high accuracy, while other technologies generate a larger number of short DNA sequences.

10. One common feature of next-generation sequencing technologies is that the cost per base of sequence generated has dropped by orders of magnitude from Sanger sequencing, and can be projected to continue to fall over the next 5 years. In 2007, 454 Life Sciences sequenced the complete genome of Dr. Jim Watson in only four months for approximately \$1 million, a dramatic reduction from the near \$3 billion and 10 years spent on Sanger sequencing for the Human Genome Project. Today, using 454 Sequencing's newest generation of technology the cost to sequence a human genome is approximately \$250,000. The cost will drop to under \$10,000 within 5 years. Although some companies currently claim to be able to sequence human genomes in that price range, the quality of the data is of questionable scientific value.

11. Advances in DNA sequencing have benefited from early stage government funding. For example, 454 Life Sciences was awarded two grants totalling \$7 million from the NHGRI for sequencing technology development. Once a technology has demonstrated technical feasibility, venture capital has been available to commercially develop the technology. 454 Life Sciences raised \$60 million in two rounds from its biotech parent, CuraGen Corporation, and outside investors. With commercial success, large life science companies have deemed next generation sequencing technology a core asset, as evidenced by the recent acquisitions of 454 Life Sciences by Roche (2007) Solexa LTD by Illumina, Inc. (2006), and a sequencing technology from Agencourt Inc by Applied Biosystems Inc. (2006). Each of these companies continues investing in the acquired technology.



1,000 Genomes Project

12. The benefit of sequencing 1,000 human genomes across known population groups is to establish the most detailed and medically useful picture to date of human genetic variation. Future study of the genetic causes of disease will benefit from a deep catalogue of human variation. This reference will allow researchers to understand which genetic differences are more likely to be normal genetic variation versus those that are causative or predisposing for disease.

13. We should not expect the 1,000 Genome Project in and of itself to yield significant scientific advances in understanding disease. That understanding will require extensive sequencing of patient samples from disease-specific populations.

14. The likelihood that these disease-specific studies will yield only predisposition genes with limited relevance to clinical practice is very small. A recent study¹ that sequenced a portion of 188 genomes from lung cancer patients confirmed and extended the number of predisposition genes, and more importantly, discovered many novel lung cancer targets, novel signalling pathways and strikingly, suggested novel treatment strategies for lung cancer from existing drugs that would not have been considered previously (Ding et al *Nature* (2008). 23:1069. Somatic Mutations Affect Key Pathways in Lung Adenocarcinoma)

15. There are many mechanisms by which diseases arise and progress that are not fully understood. The detailed analysis of many human genomes will allow for an increased understanding of the genetic basis of

- inherited genetic causes of disease – such as familial cardiomyopathies, cystic fibrosis and breast/ovarian cancers;
- the susceptibility to disease, such as cardiovascular disease, Alzheimer's, lupus, inflammatory bowel disease and adult onset diabetes, whose direct genetic linkages are poorly understood;
- somatic diseases such as cancer;
- the potential environmental causes of disease including, but not limited, to the natural microbial flora within and on each person and
- the combination of both genetic and epigenetic events that lead to diseases.

\$1000 Genome

16. The "\$1,000 dollar genome" typically refers to the costs necessary to generate the dataset of a complete, individual human genome. The total costs, including sample acquisition computer hardware costs, downstream data analysis costs, and data management will be significantly higher.

17. We will not speculate as to when it will be possible to generate for \$1,000 a complete individual human genome of sufficient quality to conduct scientific research or make clinical diagnostic decisions.

18. However, at \$1,000, human genome analysis will allow truly personalised healthcare. The benefits of improved disease management are many fold, including improving the life of the patient, increasing productivity through fewer lost workdays and a more targeted, cost-effective treatment as a result of a more comprehensive diagnosis and prescribing a safe and effective

¹ Somatic mutations affect key pathways in lung adenocarcinoma. Ding L et al. *Nature*. 2008 Oct 23 455:1069-75.



drug. A recent report estimated that the UK National Health Service spends nearly £2 billion annually to treat patients with an adverse reaction to drug treatment, which is an average of £1,824 per patient.²

19. The highly favourable pharmacoeconomics from a \$1,000 genome, and possibly even sooner at a higher price, suggest that everyone should have their genome sequenced due to the net positive societal benefit.

20. Each individual should have the right to decide if they want their genome sequenced. The data should belong to the individual and they should have the right to decide with whom to share.

Benefits to clinical practice from next generation sequencing

21. Clinical practice will benefit from next generation sequencing in the practice of truly personalised medicine. With better understanding of a person's genetic makeup, clinicians will design successful strategies for the diagnosis, treatment, monitoring and prevention of disease in a more accurate and cost effective manner. For example, drug choice may be tailored and the amount of drug or dose adjusted to an individual's genetic make-up.

22. Even before the \$1,000 genome, next generation sequencing will benefit clinical practice. One area of immediate clinical relevance is characterisation of drug resistant viral variants. Research at the Yale School of Medicine used 454 Sequencing to detect low frequency drug-resistant strains in HIV-infected patients; these resistance variants are proven to have significant impact one early stage treatment failure. Another study used 454 Sequencing to explain the emergence of drug resistant tuberculosis strains at the genomic level. Other clinical opportunities for sequencing include medical forensics, tissue transplantation, central nervous system degenerative diseases, and inflammation and metabolic diseases such as diabetes.

Funding Priorities

23. As sequencing technology is already well on the way to providing individual human genomes at an affordable price, the focus of government funding should be on developing the infrastructure to translate genomic data in societal benefit. We reiterate the recommendations submitted by Roche Applied Science on 21 April 2008.

a) The need for further investment in IT infrastructure and training

The amount of raw data that is produced by genetic laboratories will continue to grow and grow at a rapid rate, but it will need to be stored and analyzed if it is to be used effectively. This has raised key concerns within the research community:

- that investment in IT infrastructure has not kept pace with the development of technology, and, in many research laboratories, the existing infrastructure that is in place will not be capable of coping with the volume of data which will be generated; this appears to be a particular problem in clinical genetics facilities in the NHS.

- that there are no bioinformaticians within the NHS to analyze the data and the environment and a lack of IT infrastructure will hinder attempts to recruit such specialists.

² Adverse drug reactions cost NHS £2 billion. The Guardian, April 3 2008.



- that there are not enough trained individuals capable of running the IT equipment.

Unless these issues are addressed, these weaknesses will become more and more exposed. In turn, researchers will be unable to benefit from innovation in sequencing technology that is continuing apace.

b) The need for nationwide databases

A second major barrier to the translation of genomic research is the lack of a UK-wide database for storing sequencing information. Such a database would handle data from wherever it was produced and be accessible to researchers up and down the country. In the absence of this kind of resource, some laboratories have forged links independently and are sharing data, but there has been no consistency of approach.

A nationwide database would require significant investment to establish and a permanent funding stream to maintain, but the idea enjoys widespread support from across the research community.

c) Research and development funding has been focused on the 'R' at the expense of the 'D'

Over the last few years, funding for medical research has risen steadily, and will increase by a further £1.7 billion by 2010/11 under the terms of the Comprehensive Spending Review. This is welcome, but spending on development has not kept pace with spending on research – it is generally acknowledged that funding has been focused on the 'R' at the expense of the 'D'.

Again, this has hampered efforts to ensure the translation of genomic research into routine diagnostic tests, and is a crucial weakness that has yet to be addressed. The increased funding for translational research and the Biomedical Research Centres with an interest in genomics research will help to address this.

d) The lack of a clear regulatory role

The pathway for approval of new drugs in the UK is well established with the National Institute for Health and Clinical Excellence (NICE), but there is no NICE equivalent for diagnostics. The lack of clarity regarding both the regulatory and commissioning pathways presents a serious barrier to making novel molecular diagnostics available for clinical evaluation and use. However, we welcome the discussions which have recently begun between the Department of Health, industry and other key stakeholders to explore how this can be achieved, with the possibility of creating a single evaluation pathway for diagnostics.