

QUINQUENNIAL REVIEW 2021-2026

Genome Research Limited Strategic Overview



Contents

Executive Summary	1
Genome Research Limited (GRL): Mission and Vision	
Mission Vision Background to Genomes and BioData	2
Background to Wellcome, GRL and the Wellcome Genome Campus	
The Wellcome Genome Campus in 2020	3
Partnership with EMBL-European Bioinformatics Institute	
People and Culture	4
Delivery of the GRL's Mission	5
Evaluation of Impact Leadership, Governance and Management	6
Risks and Challenges	7
Funding Request and Assumptions	8
The next 25 years of the Wellcome Genome Campus	
Strategic Priorities for the next 25 years	9
An Evolving Community of Organisations with Shared Guiding Principles, Values and Kinships	10
The Wellcome Sanger Institute	
Background Mission Strategic Profile of the Sanger Institute	11
The Sanger Institute's Research Culture	
Goals, Principles and Structure of Sanger Institute Research	12
Platforms for Data Generation and Data Handling	
Incubating the Next Generation of Genome Scientists	14
Engagement with the National and International Research Community	
Achievements Between 2014 and 2019	15
Deliverables	16
Financial Allocations	17
Entrepreneurship and Innovation	
Background to Entrepreneurship and Innovation in Genomes and Biodata	
Background to Innovation on the Wellcome Genome Campus	18
Mission Expansion of the Wellcome Genome Campus and Innovation	19
Connecting Science	
Background Mission and Approach	21
Engagement Learning and Training Expansion of the Wellcome Genome Campus and the "Genome Gateway"	22

Executive Summary

Extraordinary opportunities and challenges are posed to 21st century biomedical science by the information encoded in genome sequences. Through the impact of its collective activities Genome Research Ltd (GRL) will provide global leadership in shaping and accelerating the revolution in biology, and its ramifying applications, engendered by the ever escalating wave of data from genomes.

GRL is a wholly owned subsidiary of Wellcome. The theme of all activities conducted under its auspices is “Genomes and BioData”. Its Mission is “to maximise the societal benefit of knowledge obtained from genome sequences” and incorporates three elements: Research; Innovation; Learning and Engagement. GRL has responsibility for developing and implementing strategy for the Wellcome Genome Campus, the Wellcome Sanger Institute, the Connecting Science initiative in learning and engagement, and the BioData Innovation Centre for companies operating in Genomes and BioData.

The Wellcome Genome Campus was established 25 years ago. In addition to the entities directly under GRL’s auspices (see above) it houses the EMBL-European Bioinformatics Institute, the ELIXIR hub and the Genomics England Ltd (GEL) sequencing facility operated by Illumina. Over the forthcoming 25 years GRL aims for the Genome Campus to expand physically to two and a half times its current size and to triple its workforce to ~7,000 people, increasing the activities of all three elements of its mission and the number of organisations it accommodates. This will create a world-unique environment of expertise and infrastructure in Genomes and BioData with a porous interface between fundamental research, industry, education and engagement.

The Wellcome Sanger Institute, in 2021-2026, will continue to conduct world-leading genome research, identifying new frontiers of biology to be explored through genomics. Its science will be characterised by coordinated large projects and industrial-scale data production supported by computational analyses, delivering research outputs that cannot easily be generated elsewhere. The Institute’s science will be embodied as five Programmes encompassing areas of biology and disease in which genome sequences will be transformative; Human Genetics, Cancer Ageing and Somatic Mutation, Parasites and Microbes, Cellular Genetics, and the Tree of Life. These will be supported by major data generation platforms in nucleic acid sequencing, cell culture, mutagenesis and microscopic imaging, with the Data Centre managing the outputs. The Institute will lead and participate in national and international consortial networks to maximise the impact of these endeavours and will foster the development of the next generation of genome scientists. Through its science Sanger will influence the thinking and future plans of scientists, clinicians, biomedical industries, health services, governments and publics worldwide.

The BioData Innovation Centre, in 2021-2026, will continue to provide facilities accommodating small and medium sized commercial enterprises operating under the theme of Genomes and BioData. With the Genome Campus expansion supported by our planning authority, the expectation is to substantially increase this provision over the next decade to include more, and potentially larger, companies. We will engage in contracted large-scale clinical, research and commercial sequencing, leveraging the expertise of our sequencing and informatics platforms, building further capability in the process.

Connecting Science, in 2021-2026, will continue to deliver its programmes of Courses and Conferences, markedly expanding its global training reach to more participants in low and middle income countries. There will be an increase in the scale of public engagement, on a path to conducting a national dialogue on genomics built on a robust evidence base. A bespoke new building is envisioned to house Connecting Science over the next few years which will have the material and symbolic effect of putting the public at the heart of the Wellcome Genome Campus.

Genome Research Limited (GRL)

Mission

To maximise the societal benefit of knowledge obtained from genome sequences.

Delivery of this Mission will have three elements:

■ Research

Advancing understanding of biology using genome sequences and other types of large-scale biological data.

■ Innovation

Applying genome science for human health and other societal benefits.

■ Learning and Engagement

Fostering knowledge exchange and discussion of the scientific, medical and wider implications of genomes.

Vision

The ambition of GRL is to progressively strengthen its well-established foundations in scientific research and discovery, and to build on them, developing the Wellcome Genome Campus over the forthcoming 25 years into the international centre for scientific, business, cultural and educational activities emanating from Genomes and BioData.

Background to Genomes and BioData

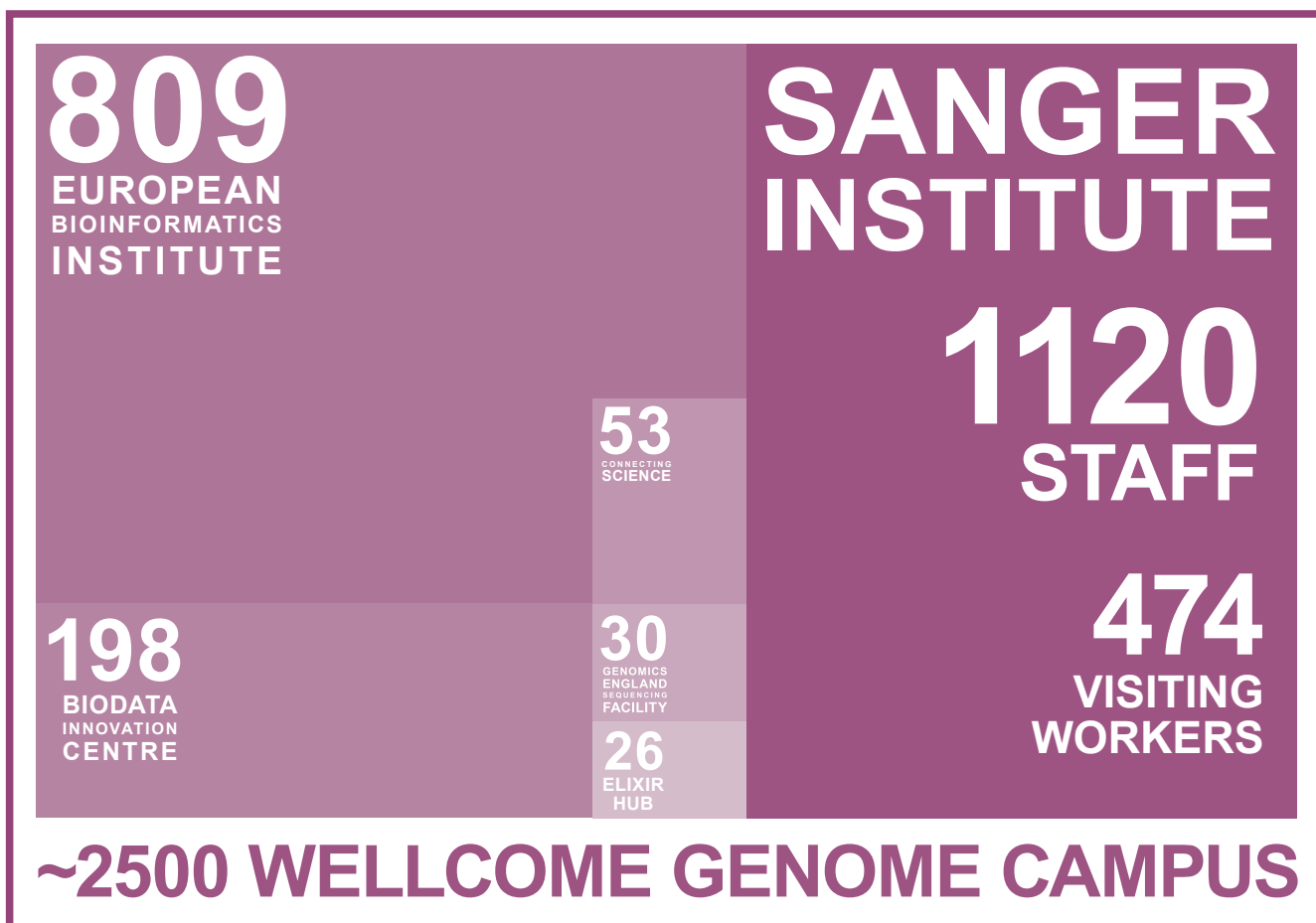
The last three decades have witnessed an extraordinary increase in the amount of data generated from biological systems. This has primarily been initiated by the advent of systematic genome sequencing, and fuelled by iterative reductions in its cost, but other sources, including biological images, are increasingly contributing. In consequence, there has been a transformation in the influence of computational approaches in biology to store, organise, analyse, share and present these data. Together, these changes are having a substantial and pervasive impact on biological science and its applications in medicine, agriculture, management of the environment and in other contexts. These changes are also engendering wider social effects with the adoption of genomics at scale by governments, health services and the pharmaceutical industry, and the creation of a biotechnology sector dedicated to generating and handling large amounts of biological data. Genomes are entering the lives of people, healthy and with disease, bringing questions and challenges pertaining to personal perspectives and public understanding.

Background to Wellcome, GRL and the Wellcome Genome Campus

To deliver its commitment to the generation of the reference human genome sequence, GRL was established by Wellcome in 1992 as a subsidiary under the auspices of which the Sanger Centre (subsequently the Sanger Institute) operated. At the same time, Wellcome purchased Hinxton Hall and its estate, and established the Wellcome Genome Campus on which the Sanger Centre was built. The EMBL-European Bioinformatics Institute (EMBL-EBI), part of the European Molecular Biology Laboratory international treaty organisation, joined the Genome Campus in 1994. Initially, GRL's responsibilities only included the Sanger Institute, while Wellcome directly managed the Wellcome Genome Campus estates, Hinxton Hall Conference Centre and the Wellcome Courses and Conferences. In 2012, Wellcome placed all these activities under GRL in order to facilitate the development of a comprehensive, long term and coherent strategy for the Wellcome Genome Campus. This strategy has been developed in close partnership with EMBL-EBI and wider consultation with academia, business, government, education and engagement organisations, and others.

The Wellcome Genome Campus in 2020

Approximately 2,500 people work on the Campus, arguably the largest single site community worldwide of expertise in Genomes and BioData. They are employed in multiple organisations. Some of these have governance and agendas independent of GRL, but all contribute to the GRL mission for the Campus and, together, constitute an aggregate of endeavour in Genomes and BioData amounting to substantially more than the sum of its individual parts. The organisations include the Sanger Institute, the EMBL-European Bioinformatics Institute, the BioData Innovation Centre (BIC) for ~10 small to medium sized companies, the Genomics England Sequencing Facility operated by Illumina, the ELIXIR hub and Connecting Science.



Current staffing figures – November 2019

Multiple substantial advances during the 2015-2020 Quinquennium have led to the current status of Campus development;

- The BioData Innovation Centre (BIC) for small to medium sized businesses opened, has accommodated nine companies and is full.
- The Genomics England sequencing facility operated by Illumina was established alongside the Sanger Institute sequencing facility.
- The new EMBL-EBI South Building was occupied, increasing the space for provision of public data services and training, and housing the headquarters of ELIXIR, the intergovernmental organisation coordinating life sciences computational resources across Europe.
- Connecting Science was created and developed into a unified learning and engagement programme with a defined mission and identity.
- Several new initiatives involving multiple entities on the Genome Campus were established including Open Targets (Sanger, EMBL-EBI and pharmaceutical companies) and the Health Data Research UK (HDR-UK) substantive site (Sanger, EMBL-EBI, University of Cambridge and Cambridge University Health Partners).

Partnership with EMBL-European Bioinformatics Institute

Sanger and GRL have a long standing partnership with EMBL-EBI. We have a shared vision of the importance of large biological data, and the openness of its provision, for the future of biomedical science. The partnership greatly enables both organisations to deliver their goals, with the data generation mission of the Sanger Institute complementing the data distribution mission of EMBL-EBI. The two Institutes collaborate extensively on data analysis, with many shared research programs across multiple biological sectors, have shared engagement and training activities, and shared people. We consider together the future direction of the Genome Campus and consult closely on national and international strategic scientific agendas.

People and Culture

The 1120 people employed by GRL include scientific faculty, other senior scientists, post-doctoral fellows, graduate students, clinicians, technicians and bioinformaticians. Our scientific workforce is supported by administration, IT, HE, health and safety, finance, legal, campus maintenance, capital projects and estates management, communications, learning, engagement and events management.

GRL is committed to the Genome Campus being a highly desirable place to work, providing our people with an inclusive and engaging environment which supports them to thrive and enables us to hire the best. We are actively and continuously improving this environment, through innovative benefits and initiatives that educate, excite and motivate. On these foundations we will continue to build an inclusive, responsible and responsive work culture, empowering delivery of our mission and addressing the needs and aspirations of our people. This culture will be clearly articulated, reasonable in its expectations, cognisant of rights and responsibilities, and owned by all.

It is a culture that:

- is built on diversity and inclusion;
- understands the different challenges encountered and types of support needed in our work environment;
- celebrates creativity, inventiveness, broad thinking and the power of the imagination;
- actively recognises different types of contribution to research, innovation, learning and engagement;

- acknowledges the competitive nature of science but emphasises the value of collaboration and manages the way we conduct research in a civilised fashion;
- expects exemplary standards of research practice;
- provides training, mentoring and career development in a range of careers;
- expects respect in interpersonal relations;
- expects collective responsibility for a safe and healthy environment.

We are committed to transparency in the way we run our organisation and to providing support for managers and leaders to embed these elements into workplace culture. Athena Swan Silver accreditation and a series of milestone plans will form a core part of our strategy to effect these aims.

Delivery of GRL's Mission

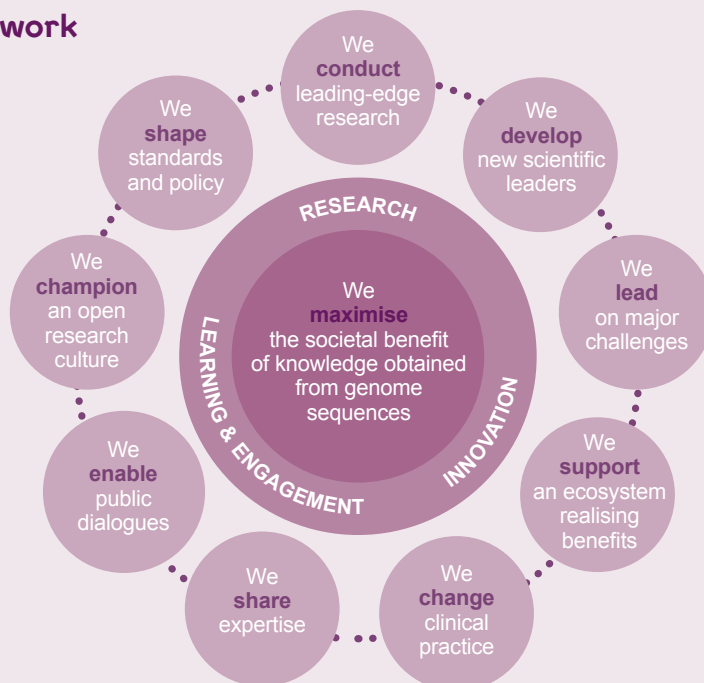
The three elements of the Mission are each delivered by multiple Campus organisations:

- | | |
|----------------------------------|--|
| Research | <ul style="list-style-type: none"> ■ Sanger Institute. ■ Society and Ethics Research Team within Connecting Science. ■ Elements of work conducted in BioData Innovation Centre companies. |
| Innovation | <ul style="list-style-type: none"> ■ Open release of Sanger Institute data, research discoveries, resources and software with direct impacts on clinical practice, public health and other applications. ■ Translation of Sanger Institute science through licensing of Intellectual Property (IP), strategic industry collaborations and externalisation of research platforms. ■ Commercial spin outs from Sanger including: <ul style="list-style-type: none"> ■ Congenica: analysis of genome data in rare diseases. ■ Microbiotica: new therapeutic strategies using the microbiome. ■ Precompetitive translational initiatives including: <ul style="list-style-type: none"> ■ Open Targets: Sanger, EMBL-EBI and multiple pharmaceutical companies using genomics to identify and validate therapeutic targets. ■ Human Data Research-UK: Sanger, EMBL-EBI, University of Cambridge and Cambridge Health Partners using genomic data and patient health records. ■ Bill and Melinda Gates Foundation Initiative on Malaria surveillance: Sanger constructing a platform providing genomics data for use by national and international public health providers. ■ Global Alliance for Genomic Health: Sanger, EMBL-EBI, ELIXIR, Genomics England, for sharing genomic data. ■ Businesses in Genomes and BioData attracted to the Wellcome Genome Campus including: <ul style="list-style-type: none"> ■ Genomics England. ■ Nine companies in the BioData Innovation Centre. |
| Learning & Engagement | <ul style="list-style-type: none"> ■ Connecting Science, including: <ul style="list-style-type: none"> ■ Scientific Conferences delivered at the Conference Centre. ■ Wellcome Advanced Courses delivered on site and around the world. ■ Public Engagement initiatives. ■ Staff at the Sanger Institute and in BioData Innovation Centre companies active through Connecting Science and other channels. |

Evaluation of Impact

To enable ourselves and others to assess the success of GRL's activities we have established the impacts we wish to achieve and be judged on. This impact framework is being developed and will serve as a basis for iterative reflection and review over the 2021-2026 quinquennium and beyond.

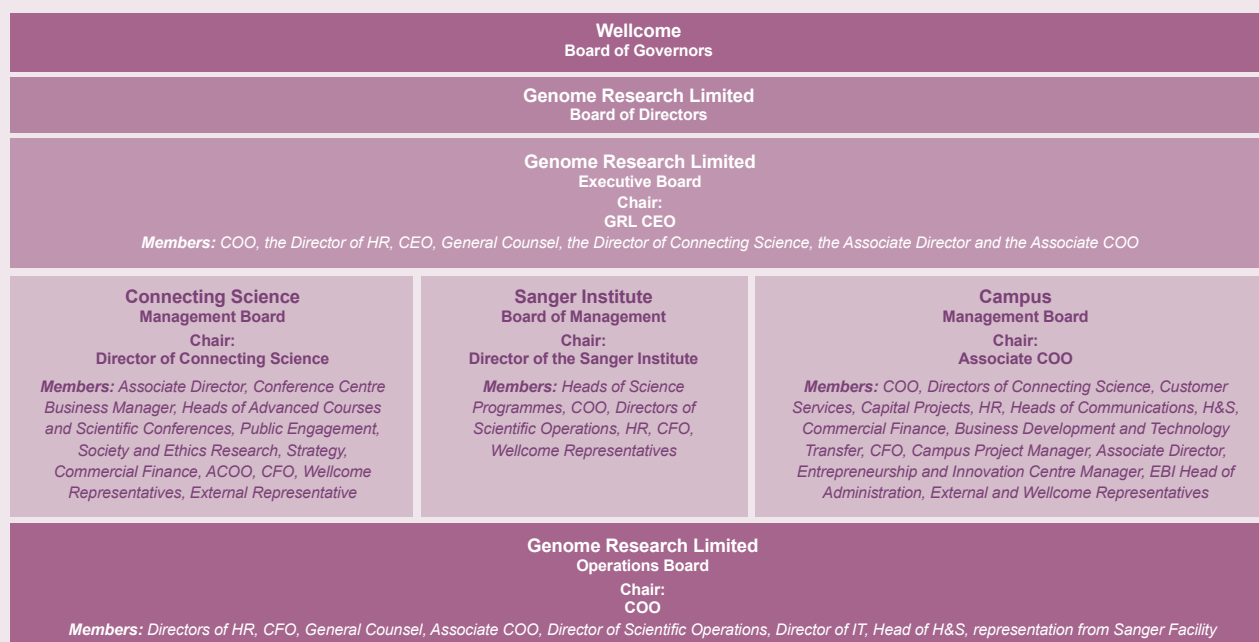
GRL impact framework



Leadership, Governance and Management

The current governance structure is represented below. GRL governance is being reviewed and there may be further changes following development of the Genome Campus expansion plan.

Current GRL Governance Structure (Job titles refer to a single individual throughout)



GRL Board is currently Chaired by the Director of Wellcome, has further representation from Wellcome, and includes others from science, industry, public engagement and government who add value to the vision, external interactions and governance of the organisation. GRL is led by its CEO (who also holds the position of the Sanger Institute Director) who, with the COO, Director of Connecting Science and GRL Executive, has responsibility for developing strategy and overseeing its implementation across all of GRL.

The Sanger Institute is led by its Director (who also holds the position of the GRL CEO) who is advised by the **Sanger Institute Board of Management (BoM)**. The Director and BoM develop science strategy and oversee its implementation.

Sanger has five **Science Programmes**, each of which is led by a Head who, together with the Institute Director, develops science strategy for the area and leads in recruitment and turnover of Faculty.

The **Faculty** conceive of and lead Sanger science. They share a commitment to understanding the biological implications and applications of genome sequences and have diverse areas of further expertise in biology, medicine, technology and computation.

There are three levels of Faculty:

- Group Leader 3 (GL3): equivalent to full professor level.
- Group Leader 2 (GL2): equivalent to associate professor level.
- Group Leader 1 (GL1): equivalent to tenure track / assistant professor level.

Sanger aims for ~35 faculty, seven in each Programme with half GL3 and half GL1/2. This structure accommodates sufficient GL3 scientists to develop long term scientific strategy in each area, sufficient GL1/2 to provide opportunities for career development and succession, and maintains adequate resources for each faculty member to conduct science at scale.

Connecting Science is led by its Director who along with the **Connecting Science Management Board** develops strategy and oversees implementation for the Advanced Courses and Scientific Conference, Public Engagement, Society and Ethics Research and the Conference Centre teams.

Campus Management Board is responsible for the development and management of the buildings and infrastructure of the Wellcome Genome Campus and for oversight of the Enterprise and Innovation element of GRL's strategy.

GRL Operations Board is chaired by the COO, who leads the provision of management and scientific operations for the Sanger Institute, Connecting Science and the Genome Campus.

Scientific Operations are led by the Director of Scientific Operations who reports to both the COO and Director. The Director of Scientific Operations has responsibility for all shared scientific facilities including the Sequencing facility, the Cell Culture Facility and the Research Support Facility (the animal house).

Management of the **Genome Campus** estates, facilities and the BioData Innovation Centre is led by the COO and Associate COO.

Risks and Challenges

We set an appropriate risk appetite across the organisation depending on function. In areas of regulatory and legal compliance we adopt stringent controls to robustly manage risk. In conducting discovery research, however, we accept that not all outcomes are achieved. Looking forward we see our major challenges coming from the areas listed below.

We will continually monitor and adapt our strategy to ensure we:

- Remain a global research leader;
- Recruit and retain our thought leaders;
- Maintain the level of scale and funding;
- Maintain our position at the leading edge of technology.

Funding Request and Assumptions

Following initial discussions with Wellcome concerning the value of our core funding award for the next quinquennium, we were advised to position the request in the vicinity of £550m. The GRL Core Funding request for the 5 year period therefore amounts to £560m. This is a substantial award that reflects the scale of the scientific endeavour and the infrastructure of the organisation. However this does represent a real terms reduction of £13m on our current funding. We will continue to leverage our core award to obtain third party grant funding for our scientific objectives and have assumed broadly comparable levels of third party between the current and next quinquennium (22%). Our projections for commercial income, however, are lower than for the current quinquennium due to the exceptionally large nature of our current UK Biobank sequencing contract. The overall income projection is therefore £830m for the quinquennial period which represents an uplift of 48% over the core funding requested.

Previous Quinquennial and future Quinquennial finances

GRL	Quinquennial Core Funding Request				All Sources Funding			
£m	Current	New	Delta	Delta >2%*	Current	New	Delta	Delta >2%*
Sanger Institute	469	506	36	(12)	739	747	8	(69)
Connecting Science	26	30	4	1	49	54	5	(0)
Campus	24	25	1	(2)	29	30	1	(2)
Total	519	560	41	(13)	817	830	13	(72)

* 1 Delta >2% represents the real terms change from assumed inflation rate of 2% pa. Positive numbers indicate growth and bracketed numbers indicate a reduction.

For the purposes of the funding request we have adopted the Wellcome inflation rate of 2% and allocated the award envelope across the various components of GRL in the same proportions as the current quinquennium. The 2% inflation is lower than our internal projection of 2.5% which is a blended rate combining staff cost inflation of 3% and other costs (excluding utilities) of 2%. Neither rate takes account of uncertainties associated with the current political environment. After a process which took account of current and planned activity levels and likely available funding, the core funding requests for the Sanger Institute and Connecting Science are £506m and £30m respectively with Campus at £25m.

It should be noted, however, that the infrastructure associated with a research campus operating at scale involves significant utility usage. Utility inflation has been running at ca.7% pa, significantly in excess of the 2% Wellcome estimated rate used in the core funding request (with a consequent estimated shortfall over the quinquennium of £1.6m). Moreover, the core funding request specifically excludes a number of infrastructure projects that will be discussed with Wellcome separately (with supporting business cases) due to their exceptional nature. The largest of these are two projects which aim to maintain the value of the campus and are refitting the Sulston Building and re-purposing the (RSF) animal house.

The next 25 years of the Wellcome Genome Campus

Our expectation is that the numbers of genome sequences generated, at the Genome Campus and elsewhere, will continue its remarkable acceleration in forthcoming decades and that the quantities of other types of biological data will similarly increase. This will have ever increasing and more diverse impacts on understanding of biology and its applications in health and other contexts.

To lead in the generation, management, extraction of knowledge from and application for societal benefit of these data, GRL has the aspiration that the Wellcome Genome Campus should grow substantially over the next 25 years. Our current expectation is that it will increase to approximately two and a half times its current physical footprint to accommodate approximately 7,000 people and include many more organisations operating under the core theme of Genomes and BioData. The trajectory to this future has been unlocked by granting of permission to plan these developments on land owned by Wellcome that is contiguous with the current Campus.

Strategic Priorities for the next 25 years

We envisage that the Sanger Institute and EMBL-EBI will remain at the heart of the expanded Campus. The two Institutes will continue to be intellectual drivers, with other academic, clinical, commercial, educational and engagement organisations attracted by their scientific missions and capabilities. Preservation of their scientific prominence and excellence will therefore be essential to achieving the wider ambitions of the Genome Campus.

We intend the future growth of the Campus to serve each of the three elements of GRL's Mission. Thus we aspire to see:

- Further not-for-profit organisations involved in fundamental or applied genome research and/or in provision of public data services.
- More businesses, including clinical genomics companies, data companies, biotechnology companies and other commercial entities ranging from start-ups to more mature businesses to facilities established by global pharmaceutical or data companies.
- A newly built facility, under the auspices of Connecting Science, dedicated to developing long term, nationally and internationally influential public engagement initiatives relating to Genomes and BioData and to science training, providing a state-of-the art permanent home for the Advanced Courses programme.
- Substantially increased accommodation for visiting and short term workers.
- Approximately 1,500 houses for permanent accommodation of individuals working on Campus and their families, together with services potentially including sports facilities, medical centre, primary school and shops, and with a conception of Campus landscape, connectivity and community looking forward to mid and late 21st century life.

An Evolving Community of Organisations with Shared Guiding Principles, Values and Kinships

We envisage the continuing close relationship with Wellcome to be a defining feature of the identity, philosophy and spirit of the future Wellcome Genome Campus. The Genome Campus and its constituent organisations will be enriched by participation in Wellcome's scientific culture and will make its own distinctive contribution to it.

Many entities attracted to the Wellcome Genome Campus in the future will not, however, be funded by Wellcome and will not be within GRL's direct responsibility and governance. These will include both commercial and not-for-profit organisations that have newly invested in the construction, development and operating of the Genome Campus. Their interests will need to be adequately and appropriately represented in a refashioned governance structure.

It will be imperative, however, for the Genome Campus to retain cognisance of its history, values, core scientific strengths, focus and global role, such that new organisations remain closely aligned to the theme of "Genomes and BioData" and to GRL's Mission. In this way, the activities in science, in business, in learning and in engagement will be mutually reinforcing through shared interests, shared expertise, shared facilities, shared infrastructure, shared outside connectivity and a pool of people committed to and capable of delivering the Mission.



The Wellcome Sanger Institute

Background

Understanding and using genome sequences is a primary impulsive force and defining attribute of 21st century biomedical science. The draft human genome was announced in June 2000. 20 years later genomics is central to exploration of biology and human disease in all its contexts and is contributing routinely to healthcare and other applications. The cost of generating nucleic acid sequences has fallen approximately one million-fold over this period and exponentially increasing amounts of sequence data are being generated with a consequent transformation in the role of computational biology. However, we are still in the foothills of extracting and using the knowledge buried in the 3,000,000,000 letters of code in the human genome and have scarcely begun to consider the rest of the genomes on Earth. Further transformative advances in our understanding and management of the living world will spring from the ever increasing accrual of human and other genomes and the myriad analytic perspectives taken. The role of the Sanger Institute is to provide intellectual, technological and operational leadership to accelerate this global scientific trajectory.

Mission

To use information from genome sequences to advance understanding of biology and improve health.

Until 2018 Sanger limited its Mission to “human and pathogen biology” and “human health”. Whilst these remain dominant themes in the 2021-2026 science strategy, the Mission has been broadened with the intention of taking a leading role in sequencing all species on Earth. Beyond its immediate impact through the activities required to deliver the “Tree of Life”, this will bring widened scientific engagement to Sanger and other entities on the Wellcome Genome Campus, extending to evolutionary biology, consideration of the Earth’s living environment and synthetic genomics. This strategic change is congruent with Wellcome’s current consideration of “Our Planet Our Health” as a new area of involvement and support.

Strategic Profile of the Sanger Institute

To deliver its Mission Sanger will:

- Be an “ideas factory” identifying new frontiers of biology and application to be explored through genome sequences.
- Be at the leading edge of development, application and implementation of genomics technologies.
- Conduct research characterised by large-scale data production that cannot easily be conducted elsewhere.
- Be expert in aggregation, analysis and interpretation of large quantities of genomic data.
- Initiate and participate in major national and international collaborative research initiatives.
- Incubate the next generation of scientists and clinicians in genome research.

The Sanger Institute's Research Culture

Sanger is a genome research institute. In developing plans for 2021-2026 we reconsidered our Mission and strategic principles and reaffirmed our belief in genome research conducted by exceptional scientists conceiving new scientific directions in an intellectual environment shaped by the opportunity for large-scale data production and analysis. Our science is often systematic, globally descriptive and foundational resulting in reference data sets to be used by the wider scientific community. With generous core funding enabling agility and independence in initiating new directions we aspire to conduct research which cannot easily be done by other research organisations that lack our assemblage of scientific and technical expertise, supporting infrastructure and organisational experience. We continue to strongly champion open release of data and global collaboration, scientific cultural commitments which flow from our organisational origins and traditions.

Our thought leaders were mandated to be imaginative and adventurous, considering scientific agendas and technologies that might be deployed towards the end of the quinquennium in 2026 as well as experiments feasible at its inception in 2021. Developing new threads of genome research and leaving old ones behind, taking intellectual and technological risks, being driven both by elemental curiosity and potential application, considering long and short term impact, being undaunted (indeed, being inspired) by challenges of scale and conceiving a limited number of major initiatives involving the collective scientific endeavours of multiple Faculty, multiple Programmes and multiple organisations have all been central to the philosophy in developing the scientific agenda for the next five years.

The portfolio of science described is an indicative outline of the scientific directions the Sanger Institute will take in 2021-2026. As in the past, however, changes to our science plans may occur during this period and may entail changes in Faculty, in Programmes and in Platforms. Technology innovation is rapid in genomics and can lead to major shifts in research direction. A key strength of the Institute is its ability to be flexible and adaptable, adopting new lines of enquiry quickly, and reallocating resources in order to do so.

Goals, Principles and Structure of Sanger Institute Research

There are three major goals of our science:

- To provide new insights into the normal biology of humans and other organisms.
- To enhance understanding of the pathogenesis of human disease and provide the scientific basis for improving disease prevention and management.
- To elucidate the evolutionary tree of life on Earth and advance the scientific knowledge base for management of the living environment and for synthetic genomics.

To achieve these goals we will adopt three major experimental approaches:

- Analysis of DNA sequence differences between individual genomes of the same species. The sequence variation studied will be naturally occurring and engineered, inherited and somatically acquired and include differences between the genomes of individual humans, of individual isolates of disease-causing microbes or their vectors, and of individual cells. These studies will provide insights into the causal roles of genes in healthy and disease traits, into the histories of humans, microbes and cells and into the causes of sequence variation.
- Analysis of gene expression and epigenomic features of the genome at the single cell level to discover and classify cell types in health and disease and to understand their developmental and functional relationships.

- Analysis of reference genomes across the Tree of Life to understand evolution, provide insights into gene function and to generate informational resources for monitoring of the living environment and synthetic biology.

These scientific approaches will depend on a limited set of core technologies into which we will invest substantially, proactively conducting R+D to be at the leading edge of implementation, and which will be optimized for delivery at large-scale. These will include:

- Nucleic acid sequencing.
- Mutagenesis.
- Cell culture.
- Microscopic and spatial imaging of nucleic acid profiles of cells in tissues.
- Computational analysis of the data volumes generated.

To conceive and orchestrate Sanger science there will be five scientific Programmes, each encompassing a major sector of biology or disease to which genomics will make a transformative contribution.

- Five Programmes is judged as representing the appropriate balance between, on the one hand, maximising breadth of scientific impact and, on the other, concentration of resources to achieve science-at-scale.
- Each Programme will limit itself to two overarching Projects. This Project based structure reflects the Institute's history of conceiving and delivering substantial and strategic scientific initiatives with a beginning, a middle and an end, and its continued embrace of the attendant culture of coordinated science and collective endeavour conducted at scale.

The **Cancer, Ageing and Somatic Mutation Programme** will use genome sequencing and experimental models to study the causes and consequences of somatic mutations through the course of life, across organ systems, in health and in disease.

The **Cellular Genetics Programme** will use transcriptome sequencing and microscopic spatial tissue imaging to redefine the repertoire of human cell types, and to understand their functional states and relationships in health and disease, during development and adulthood.

The **Human Genetics Programme** will use genome sequencing and experimental cell models to understand the genetic causes and biological mechanisms of disease susceptibility and progression, focusing on developmental disorders and diseases of the blood and immune system, transforming the clinical utility of human genetic variation.

The **Parasites and Microbes Programme** will use genome sequencing to generate large longitudinal datasets over time, transforming understanding of the evolutionary dynamics of disease-causing bacteria, protozoa, helminths, mosquitoes and the human microbiome with application in pathogen surveillance for disease control.

The **Tree of Life Programme** will use new sequencing technologies to produce high quality reference genomes across the diversity of eukaryotic life, analysing the data to understand life's origins, conserve biodiversity and provide the underpinnings of a new biotechnology.

Platforms for Data Generation and Data Handling

Sanger's large-scale data production and data handling platforms provide the foundations of our scientific strategy. Sanger will have three major data production platforms in 2021-2026.

- The sequencing facility will deliver standard and bespoke sequencing products based on a wide range of preparative protocols delivered on machines from multiple technology providers. R+D of new technologies, particularly to deliver data at scale, will be a major focus. The products will include short and long DNA sequence reads; DNA sequences from small amounts of DNA, from damaged DNA, from different species, from single cells and from single molecules (duplex sequencing); RNA sequences from single cells; and sequencing capturing various forms of epigenomic feature. We will seek large-scale commercial sequencing contracts that will both leverage and enhance the scale and capabilities of our platforms.
- The Cell Generation and Phenotyping facility will derive, maintain and experimentally manipulate human normal tissue and cancer derived immortal cell lines and organoids at scale, enabling implementation of large-scale small molecule and CRISPR mutagenesis screens.
- Microscopic, spatial imaging of tissues delivering *in situ* nucleic sequence data from normal and diseased human organs will be the major new departure in our technology repertoire during 2021-2026, enabling new approaches to analysis of tissue architecture and function.
- In the next quinquennium, we will more closely integrate data, informatics and IT. To help drive this change we will create a senior role to lead in the coordination and delivery across the organisation. Through this function, we will adopt an Institute wide review of our approach to data, informatics and the underlying IT infrastructure. Working with our partner organisation, EMBL-EBI, there are significant benefits to be realised in harmonising our approach to data and informatics, including compression software, and developing the right tools for our researchers so they can achieve their scientific and translational goals.

Incubating the Next Generation of Genome Scientists

This is a central strategic component of our mission which will be delivered in multiple ways:

- The Graduate Programme will qualify ~15 new Doctoral scientists each year. There is no limitation on country of origin, diversity in recruitment is a core goal.
- Clinicians with expertise in genome science are important future leaders as genomics becomes a common feature of healthcare. We will qualify approximately three clinicians every year as PhDs, hosting additional individuals in Post-Doctoral Fellowships, Career Development Fellowships, and Group Leader 1/2 Faculty positions.
- It can be challenging for talented individuals from low and middle income countries (LMIC) to be accepted onto highly competitive graduate programmes in high income countries. We will therefore continue our MPhil programme for three students each year from LMIC which provides the experience and credibility for successful application to such programmes.
- At any time there are more than 100 Post-Doctoral Scientists at the Institute. We will address their specific concerns and needs through a dedicated programme organising retreats, mentoring opportunities, training, industrial interactions, career development and career advice, considering pathways within and without academia.

Engagement with the National and International Scientific Research Community

Sanger will be an organisation with outward-looking scientific engagement operating at the centre of a network of science. This is a core and long-standing cultural theme extending beyond standard investigator-led collaboration to an inclusive organisational penumbra fostering greater collective intellect. Through it we aim to both empower genome research conducted by others and to enrich Sanger's scientific portfolio.

We will achieve scientific engagement in the next five years through the following activities:

- Our Associate Faculty, Honorary Faculty, International Fellow, joint Post-Doctoral Fellow programmes and Joint Appointments with other organisations.
- Strategic relationships with the EMBL-European Bioinformatics Institute and the University of Cambridge.
- Partnership with the Broad Institute in multiple research endeavours.
- National and international collaborative research initiatives including the Human Cell Atlas, the Darwin Tree of Life Project, the Common Disease Consortium, the Human Cancer Model Initiative, the ICGC/TCGA Pan Cancer Analysis of Whole Genomes (PCAWG) and multiple other research consortia.
- Open Targets, a consortium of the Sanger Institute, EMBL-EBI and pharmaceutical companies including GSK, Takeda, Biogen, Celgene and Sanofi-Pasteur, dedicated to identifying and validating new therapeutic targets through genome variation based approaches.
- Health Data Research UK (HDR-UK) which aims to use UK National Health Service patient data for health research and for which Sanger, EMBL-EBI, University of Cambridge and Cambridge University Health Partners together constitute one of six national sites.
- Participation in large-scale DNA sequencing provision, including the Vanguard Project of 50,000 whole human genomes for UK Biobank and subsequently a further contribution of 225,000 genomes to complete the sequencing of UK Biobank (500,000 samples in total).
- Collaboration with Genomics England.
- Hundreds of investigator-led collaborations.

Achievements

The Institute takes a broad perspective on its achievements, considering its outputs in scientific publications and data resources, its impact on training and career development, its translation of genome science from theory to practice, its development of spin out companies with the potential for global impact, its wider leadership role, its influence on policy and also the changes to our organisation we have wrought to maintain scientific momentum and create an excellent working environment for our people.

Between 2014 and 2019

- Sanger Institute personnel were authors on 3368 scientific research publications in the six years from January 2014 to December 2019. These range from studies in which we had primary leadership in conception and delivery to studies in which we made a relatively minor contribution. To highlight publications in which we have provided leadership, those including Sanger authors in first and/or last authorship positions have been identified, of which there were 819 since the beginning of 2015.
- Of these 3368 publications, 2868 were published over the period of five years between January 2014 and December 2018 and obtained 124,911 citations by December 2019. 36% (n=1020) of these publications were in the top 5% of cited papers for the period considered, and 16% (n=446) in the top 1%.

- For the past five years the Sanger Institute had on average 120 Post Doctoral Fellows (PDFs) each year, and through this programme 446 PDFs and Career Development Fellows have been inculcated with Sanger science, culture and ethos since 2015.
- One new company, Microbiotica, was launched with an initial investment of £8m progressing to a strategic relationship with Genentech valued at US\$534m. The Sanger spin out Congenica was recognised as 2019 UK MedTech company of the year and serves as the sole clinical decision support partner to Genomics England.
- We contributed to and shaped the UK Life Sciences Industrial Strategy, national genomics strategy, reports from the UK Chief Medical Officer and international health organisations, the BioIndustry Association Genomics Advisory Group, and provided evidence on future applications of genomics as well as freedom of movement for scientists.
- £96m of third party funding beyond the core Wellcome award was obtained, increasing the Institute's total revenue by 21%.
- Leadership roles in major consortial initiatives including the Pan Cancer Analysis of Whole Genomes, Human Cell Atlas, Human Cancer Model Initiative, Deciphering Developmental Disorders Project and the Darwin Tree of Life Project were taken by Institute scientists.
- Our science strategy, and the infrastructure required to deliver it, have been reviewed and reshaped. A new science Programme in the Tree of Life has been established, the Computational Genomics programme closed and the Infection Genomics and Malaria programmes merged to form a single Parasites and Microbes programme. With diminishing requirements for mice we made the decision to close the Animal Facility and initiated the process to effect this.
- We have reviewed and evolved the way we run the Institute considering our Governance, our Culture, our effectiveness in Equality, Diversity and Inclusion, the strength of our Research Practice and our Faculty model.

Sanger Institute Deliverables 2021-2026

The scientific outcomes and impact of the science we conduct will be assessed through the impact framework outlined. To provide a high level impression of our activity during this period we have summarised the data sets we aim to generate. **In 2021-26, we are aiming for:**



950

cell lines derived



2,500

CRISPR screens



2,000

human genomes for germline
human sequence variation



14,000

de novo genomes
sequenced



20,000

human genomes
for somatic human
sequence variation



30,000

human single cell
transcriptomes



50,000

plasmodium
genomes



250,000

microbial
genomes

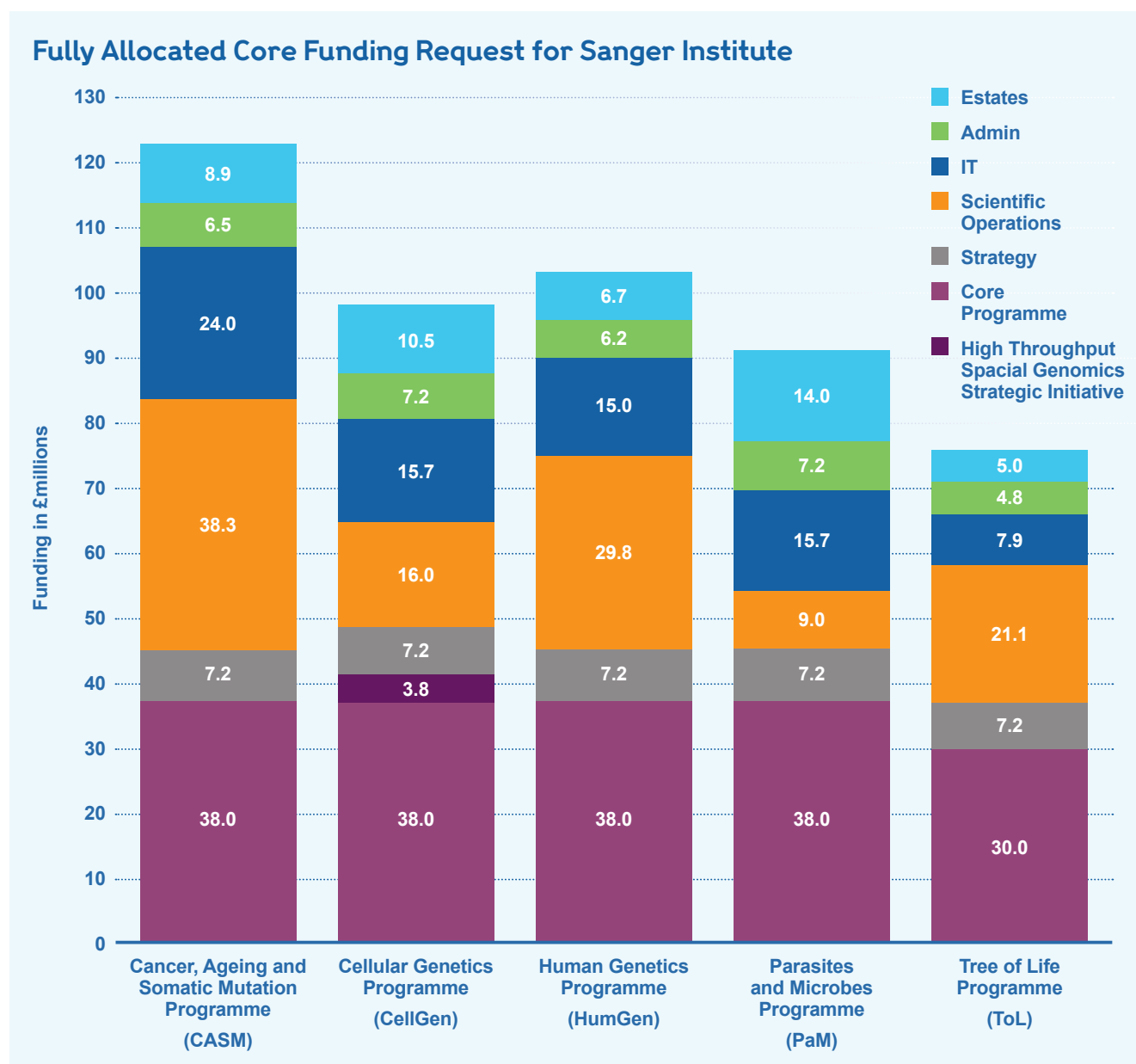


400,000

mosquito targeted
amplicons

Financial Allocations

Excluding the Open Targets and Associate Research Programme (£15m) the core funding request for the next quinquennium is £491m for the Sanger Institute.



All the costs of the Institute are incurred in order to achieve the Institute strategy so can be linked back directly to the scientific programmes. This enables a clearer representation of the resources allocated to each programme which is demonstrated in the fully allocated chart above. The initial allocation to the five programmes of £182m is increased by their respective shares of platform resource utilisation (based on demand projections including historic data storage for IT) and support services (largely headcount and footprint based).

At circa £124m the CASM costing reflects both the established nature of the programme and the costs of sequencing and IT associated with the programme. The CellGen allocation reflects the relatively high costs of imaging and the HumGen allocation reflects high historic IT storage costs. PaM is the only programme where IT platform costs exceed the SciOps costs reflecting the high level of compute resource needed for relatively small size samples in sequencing. The Tree of Life programme is being established over the period and the resource allocation assumes an ability to use IT and new sequencing technologies efficiently.

Entrepreneurship and Innovation

Background to Entrepreneurship and Innovation in Genomes and BioData

Discoveries made through fundamental research on genome sequences have provided a multitude of opportunities for application. Genome science has spawned a major industry generating, managing and interpreting sequence data which already makes a substantial contribution to the delivery of societal benefits, for improvement of human health and in other contexts. These opportunities are reflected in the 2017 UK Life Sciences Industrial Strategy placing “Genomics in Medicine” at its centre and in the anticipation that the global genomics market will be at \$36billion by 2024. We anticipate that opportunities for application and innovation will continue to grow in the future.

Background to Innovation on the Wellcome Genome Campus

Supporting these opportunities to flourish entails the availability of particular expertise sets and infrastructure. The 2500 people and the platforms in sequencing and data management on the Wellcome Genome Campus constitute a globally unique aggregation of scientific and technical skills in Genomes and BioData. This skill pool provides robust foundations for application and innovation by organisations already on Campus and is the major feature attracting other organisations working in Genomes and BioData to Campus.

Recognising this opportunity, Wellcome funded the building of the BioData Innovation Centre to house companies spun out of Sanger and EMBL-EBI and to accommodate others which would benefit from, and contribute to, the collective intellect, infrastructure and collaborative culture of the Wellcome Genome Campus. The BioData Innovation Centre opened in 2016, quickly became full and, in 2019, accommodates nine companies that form a vibrant community fully embedded in wider Campus life. Wellcome also funded a building which houses the Genomics England sequencing facility. This has generated clinical sequencing data for the UK government’s 100,000 genome project. Thus the Genome Campus is making a substantial contribution to clinical sequencing provision in the UK.

From Sanger’s science in the 2015-2020 Quinquennium one new spin out company was created, Microbiotica, which will develop new therapeutic approaches to disease management using the microbiome. During this period, Sanger also established itself as a provider of large-scale DNA sequencing services undertaking the sequencing of 275,000 whole genome sequences (out of a total of 500,000) for UK BioBank.

Mission

To deliver Innovation based on Genomes and BioData to improve healthcare and bring wider societal benefits.

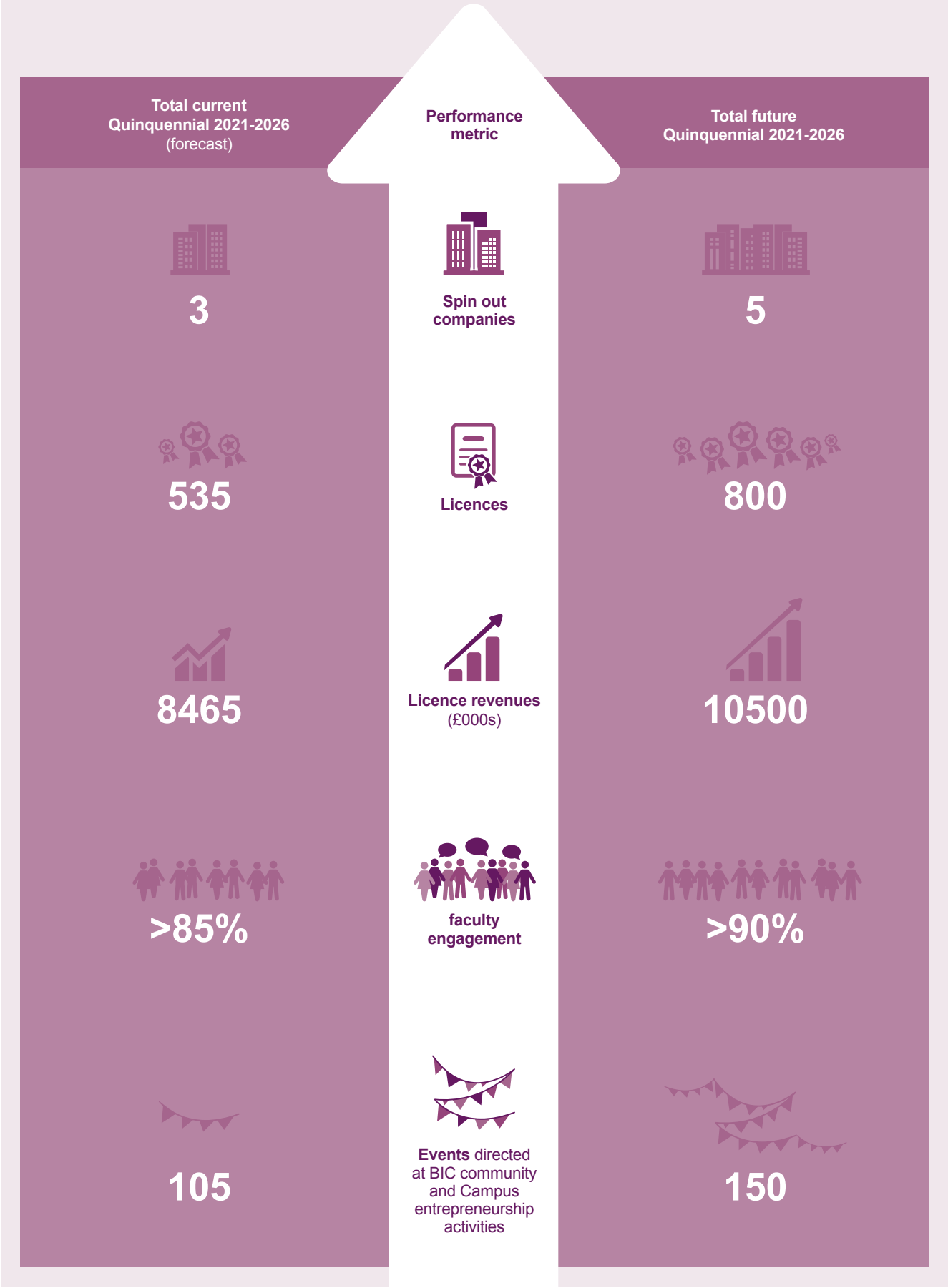
GRL will deliver this mission by:

- Creating a unique ecosystem on the Wellcome Genome Campus in which application of Genomes and BioData research is fostered and supported.
- Attracting not-for-profit and commercial enterprises innovating in Genomes and BioData to establish themselves on Campus. Applying our policy for new organisations which requires core activity in the field of Genomes and BioData will maintain and strengthen the Campus theme.
- Launching an on-site (Pre) Accelerator programme to deliver entrepreneurial support for the development of talent and early stage opportunities.
- Enabling translation of Sanger Institute scientific discoveries by public release of data and knowledge, licensing IP, strategic collaborations, and establishment of spinout companies.
- Providing services in large-scale contract genome sequencing for clinical and research purposes.
- Actively establishing synergistic partnerships with external organisations.
- Actively informing and encouraging interest from Venture Capital.
- Building programmes of activities involving all on Campus for developing translational awareness and enabling new enterprises, inspiring and developing potential entrepreneurs
- Sharing our knowledge of innovation in the genomics sector through active involvement in local, national and international initiatives in the field.
- Approaching technology translation as a deliverer of societal benefit ahead of financial return.

Expansion of the Wellcome Genome Campus and Innovation

In 2019 we had a resolution to support our planned extension of the Genome Campus from our planning authority which will provide a new and extraordinary opportunity for innovation over the next 25 years. We will explore erecting temporary buildings to satisfy immediate needs for new companies and, in the longer term, will attract investment to provide permanent space for a range of sizes of enterprises, including more SMEs, “follow-on” space for companies which have grown from SMEs and for larger, established commercial partners. This space will likely include both “dry” and “wet” laboratories.

Entrepreneurship and Innovation performance metrics



Connecting Science

Background

Just as Genomes and BioData are having far reaching implications across biomedical research, they are also being integrated into culture and society. Genomics has become part of major life decisions affecting individuals and their wider families; the direct-to-consumer testing market is taking access to genetic information out of the hands of professionals and placing it directly in the hands of the public. Concepts of genomes and inheritance are changing our understanding of ourselves as individuals and humans, as well as our place in the world around us. As genomes become ever more pervasive across all aspects of our lives, the third element of GRL's strategic plan to "foster knowledge exchange and discussion of the scientific, medical and wider applications and implications of genomes" becomes ever more pertinent.

Mission and Approach

Connecting Science, GRL's learning and engagement programme, will enable everyone to explore the impact of genome science on health, research and society.

Its strategic goals are in two thematic areas:

- Learning and training for professional audiences.
- Engagement and social science research involving public audiences.

It will deliver its Mission in these two areas through the activities of four operational groups:

- The Wellcome Advanced Courses and Scientific Conferences which develop and deliver training and conferences in genomics and related disciplines to support knowledge sharing and enhance careers.
- The Wellcome Genome Campus Conference Centre which provides a high quality events venue supporting delivery across Connecting Science activities.
- The Public Engagement team which facilitates dialogue between Genome Campus researchers and the broader public to enable everyone to be engaged with present and future directions of genome research.
- The Society and Ethics Research Team which undertakes quantitative and qualitative research into public attitudes to genomics and data sharing that informs policy and practice.

Engagement

The increasing accessibility of genomic information compels us to take a step-change in the scale of our public engagement work, and we will support dialogues around this topic on a truly national level. Strongly rooted in social science research on public attitudes to, and awareness of, genomics, we will create a robust evidence base that will inform all our engagement activities. Public audiences will be at the heart of this work, and we will partner with diverse community groups to ensure that all voices are heard in an equal and inclusive manner, working collaboratively to create engagement outputs that genuinely resonate with different public audiences. We will continue to work with Sanger Institute research programmes to facilitate opportunities for engagement with public audiences, including working with community and patient groups to contribute their perspectives to research projects. Key initiatives will be developed in partnership with all five research programmes.

Learning and Training

We will match the global impact of the Sanger research programmes by increasing our commitment to democratise access to genomics knowledge and skills through training provision in low and middle-income countries, including experimenting with novel delivery methods. Capacity building and skills development in these nations will ensure that their populations are able to benefit from the results of genomic research, as much as those in the global north. Building on our current activities, we will expand our reach by growing our online training offer and integrating it more deeply with our face-to-face courses. We will also deploy innovative new remote classroom models, allowing content to be taught simultaneously in multiple sites across a region or even continent.

Genomics has recently become part of the school curriculum in the UK, and we will enhance our current schools programme to recognise this. We will create a programme based around specific touchpoints in the curriculum, integrated with work placement opportunities in the Sanger Institute. The programme will include regular interactions between students and Institute staff to showcase the diversity of available career paths, and contribute to core GRL Equality, Diversity and Inclusion goals through our 'Science For Everyone' programme aimed at addressing unconscious bias within science teaching.

Expansion of the Wellcome Genome Campus and the "Genome Gateway"

The approved expansion of the Wellcome Genome Campus presents the opportunity to underpin Connecting Science's ambitions with a new purpose-built facility supporting and enhancing the delivery of our work programme, a project with the working title of "Genome Gateway". The "Genome Gateway" would provide a permanent home for Connecting Science, with a bespoke conception and design, fostering long term and strategic thinking at a national and international level about Learning and Engagement in Genomes and BioData. It is envisioned to be physically at the centre of the Campus putting the public symbolically and materially at the core of GRL's vision. The expansion of the Campus is just one indicator of the growth of genomic knowledge and technologies, and our five year plan is based on ensuring that our ambitious learning and engagement agenda matches this pace of growth.

Connecting Science Facts and Figures 2014 - 2019



136

training courses
held on Campus with



36

training courses
in LMICs with



90

conferences
with

3,395 
participants

823 
participants

12,210 
delegates



10

online courses
developed and
delivered with

11,952 
active learners



737

teachers
participating in
CPD programmes



166

new STEM
Ambassadors
recruited



17,835

student
interactions
with education
programmes



50,534

face-to-face public
engagement interactions
supported (on and off)
Campus



3,743

visitors to
Campus for
Open Saturday



45

peer-reviewed
publications



50,000

research survey
public participants



87,248

Conference Centre
delegates

