# Research Vision

Genomics-enabled research to drive healthcare for the benefit of everyone

Genomics England

# Contents

# Foreword

## Professor Matt Brown

Chief Scientific Officer,
Genomics England

In 2013, we embarked on a remarkable journey with the launch of the 100,000 Genomes Project and we recently celebrated Genomics England's 10-year anniversary. The current year also marks the 75th anniversary of the NHS, a testament to the enduring spirit of public health and dedication of so many who contribute to the provision of healthcare in the UK.

As we recognise these important milestones of both the NHS and Genomics England, we find ourselves at a pivotal point. The progress of the past decade, catalysed by the 100,000 Genomes Project, scientific breakthroughs, technological advancements and an unwavering commitment to genomics, has paved the way for a future of unprecedented possibilities in research and healthcare. I am delighted to share with you our Research Vision, which reflects on where we are now and sets out the course for enabling a greater diversity of transformative research over the coming years.

While Genomics England has taken on new challenges, whole genome sequencing (WGS) for rare conditions and cancer will remain at the heart of what we do. Our commitment to 100,000 Genomes Project participants is unwavering, acknowledging their role in advancing science for generations, and we recognise there are still those without diagnoses. Venturing into new genomic research areas, we aim to enhance the impact for these participants and for NHS patients.

The launch of the Generation Study shows our dedication to early diagnosis and intervention, investigating how the power of genomics shortens the diagnostic odyssey for people with rare conditions. Our commitment to diverse data acquisition demonstrates our understanding that genomic insights must be representative of our multi-ethnic population, guaranteeing equitable healthcare for all. Additionally, our Cancer 2.0 programme signifies our dedicated efforts to refine cancer intervention, leveraging new genomic technologies to usher in an era of precision cancer care. These programmes are building the foundations for the future of genomic healthcare, fuelled by research and innovation.

Multimodal data, connecting imaging and genomic, transcriptomic and other 'omic data will increasingly hold clinical relevance for the diagnosis and treatment of disease. Facilitating researchers and clinicians in this field is vital, as is the need for us to build expertise in various analytical approaches and in applying artificial intelligence (AI).

Trust and ethics have always been a cornerstone of our operation and will only gain in importance as we move to a truly data-driven learning healthcare system. Central to this is the ethos of inclusion and collaboration. Participants and patients remain at the heart of our endeavours and are the driving force behind our discoveries. Their active involvement not only empowers them but also propels us forward and ensures that our research remains grounded in real-world impact.

Our partnership with the NHS to deliver the national WGS clinical service continues to be pivotal – a true embodiment of the synergy between cutting-edge research and frontline clinical care. Many other major UK and international genomics programmes have been initiated or are reaching maturity, creating more opportunities for collaboration. Building our international connections and welcoming greater numbers of international researchers to use our Research Environment are key parts of our Research Vision.[1] With this, Genomics England commits to wide data accessibility, aiding global research and progress in genomics for public benefit.

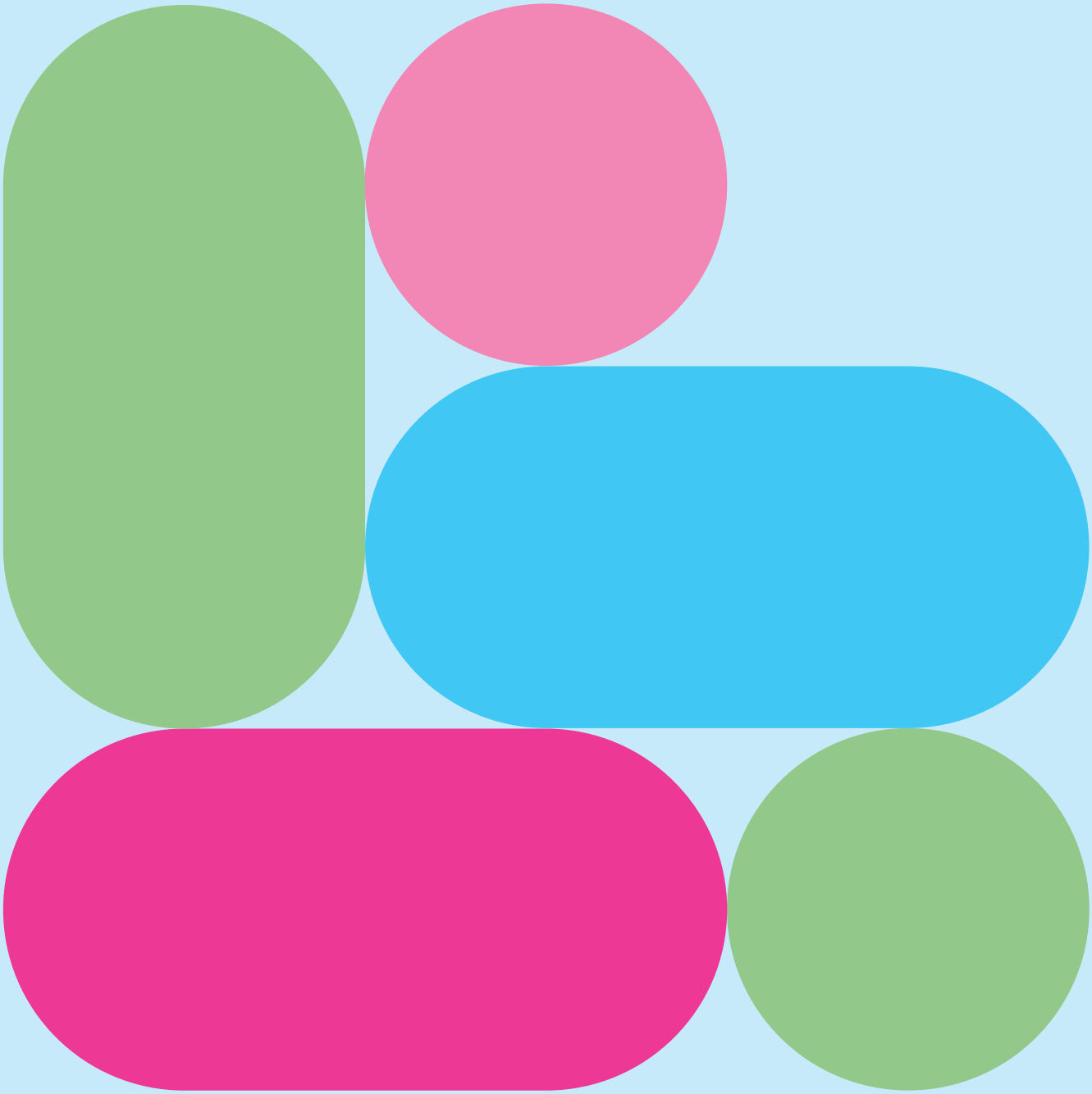
I extend my heartfelt gratitude to the participants, researchers, clinicians and partners who have made the past decade an extraordinary chapter in the annals of genomics. With your help, in the past year alone, over 1,500 potential diagnoses were made in our Research Environment and have been returned to NHS clinical teams through the Diagnostic Discovery pathway. Each of these discoveries is a testament to the potential that genomics holds in unravelling the complexity of disease, informing treatment decisions and, ultimately, transforming lives.

By continuing to work together to expand our knowledge and the application of genomic, multimodal and multi-omic data, we will remain at the forefront of genomic innovation and pave the way for a future of personalised and precision medicine for the benefit of everyone.

“With your help, in the past year alone, over 1,500 potential diagnoses were made in our Research Environment and have been returned to NHS clinical teams through the Diagnostic Discovery pathway.”

# Our vision

Over the past decade, genomics has demonstrated its potential across both research and healthcare to deliver huge advancements in the understanding and treatment of conditions within many disease areas. To accelerate this progress, we provide an outstanding data offering and researcher experience alongside our clinical services. By supporting global use of genomic and health data collected across the UK healthcare and research ecosystem, while ensuring these data remain secure under our stewardship, we can inform fundamental science, new diagnostics and therapies and innovative technologies, as well as transform clinical pathways. We believe that the research we enable should be done everywhere, safely and securely, for the benefit of everyone.



# Where we are now

When Genomics England was established in 2013, our horizons and goals were shaped by the technological capabilities of the time. We focused on the development and validation of DNA WGS approaches for rare conditions and cancer, including research into diagnostic and therapeutic approaches. Since then, there have been major advances in science, technology, analytics and our understanding of the complexity and diversity of genetic conditions, including:

- The development of new technologies enabling large-scale assays in proteomics, RNA sequencing, immune profiling, long-read sequencing and epigenetic sequencing, as well as tissue-level and single-cell approaches
- Novel research insights translating into clinical applications of polygenic risk scores and pharmacogenomics
- The emergence of extensive, publicly available bioinformatic resources across the spectrum of 'omics approaches, enabling large-scale multi-omic and multimodal research
- Recognition of, and responses to, the need to advance the involvement of non-European ancestry groups in genomic research
- Advances in analytical methods, AI and cloud computing, enabling the investigation of big data challenges
- Innovations in therapy development for genetic conditions (e.g. gene editing and antisense oligonucleotide therapies) and cancer (e.g. personalised cancer vaccines)

Years of significant and ongoing government investment in world leading healthcare and research initiatives have positioned the UK as a leader in the provision of genomic healthcare. This position is driven forward by the NHS infrastructure, including the NHS Genomic Medicine Service (GMS) – a national service that provides access to genomic testing for patients with specific conditions.

As an organisation, we have undergone a major transformation. We are now a core partner for WGS-based diagnostic services in the NHS, bringing data from the NHS GMS into our Research Environment. We have expanded the community of researchers who engage with us and the diversity of projects we support as well as establishing long-term academic, life sciences industry and strategic partnerships to deliver greater impact for participants and patients.

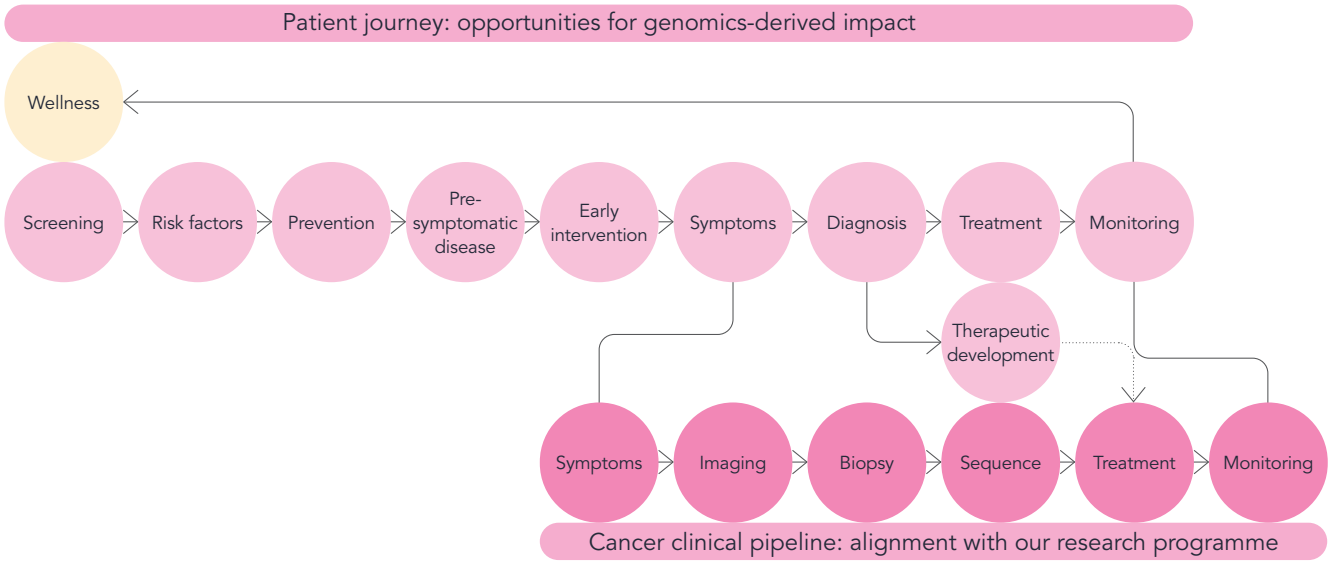
This research vision aligns with the shared aspirations of our partners, including government and the NHS, as set out in statements and strategies, including Genome UK (2020), the Life Sciences Vision (2021), the NHS data strategy (2022) and the NHS genomics strategy (2022).[2-5] It brings together our different strands of research activity and sets out what we will strive for over the next two years in order to facilitate the best use of health data for research across the UK genomics ecosystem and beyond.

# Areas of focus

Given the legacy of the 100,000 Genomes Project and our central role in the provision of WGS in collaboration with the NHS, we have a major focus on rare condition and cancer genomics. We forecast that by 2026 the number of participants who volunteered their data for research, with linked clinical data including for common diseases, will have increased considerably. This will make the National Genomic Research Library (NGRL) one of the world's largest datasets for rare conditions and cancer research as well as for common disease research. Where possible, we will support and develop approaches in common disease research and clinical implementation of genomics; however, our immediate focus remains on the development and validation of novel methods to improve genomics-informed diagnosis and therapies for cancer and rare conditions.

## Cancer

Considering clinical aspects of cancer, our priorities will be to better enable and validate next-generation sequencing for cancer. Delivered in partnership with the NHS and the National Pathology Imaging Co-operative, our Cancer 2.0 programme evaluates cutting-edge genomic sequencing technology to improve the accuracy and speed of cancer diagnosis and the use of AI to analyse a person's DNA, alongside other information such as routine scans to drive novel research discoveries. The cancer pipeline from diagnosis through to treatment, how it aligns with our research programme and where it sits within the broader patient journey are outlined below.




We will drive research to advance this process by:

- studying alternative methods to fresh frozen sample preservation to enable an easier sample handling pipeline
- investigating tumour-only analytic methods to determine if normal germline sequencing is still required to speed up turnaround times
- assessing the value of using WGS data in the development of personalised cancer vaccines
- investigating the performance of locally deployed long-read sequencing with the aim of reducing turnaround time for sequencing of time-sensitive cancers such as acute leukaemias
- increasing the insights generated from genomic sequencing by:
    - investigating long-read and epigenetic sequencing
    - studying novel decision support approaches, including the use of AI
    - researching the value of multi-omic and multimodal approaches to personalised cancer medicine
- enabling investigations into the value of WGS data for finding potential participants for enrolment into clinical trials
- continuing to build the evidence base for sequencing, including by health-economic assessment

“Cancer 2.0 aims to explore innovative genomic technologies that can have a positive impact on cancer patients. This includes long-read sequencing for rapid, precision diagnostics, treatment planning, and the multimodal data project to enhance predictive models for diagnosis, prognosis, and treatment response.”

**Cancer 2.0 programme**
Emma McCargow,
Programme Lead - Cancer, Genomics England

“In collaboration with others, including the NHS, BioNTech and Moderna, Genomics England has contributed to the development of pathways for clinical trials of personalised cancer vaccines using mRNA vaccine technology. We are particularly focused on developing and validating improved laboratory and analytical approaches to cancer profiling to reduce the time to deliver data for vaccine design and improve coverage of critical cancer neo-antigens. mRNA vaccines also have potential for the treatment of rare genetic conditions. We are developing pathways to better support research, including clinical trials of their use, such as by better linking patients with rare conditions and genetic diagnoses through our 100,000 Genomes Project or partnership with the NHS GMS with relevant clinical trials.”

**mRNA vaccines**
Professor Matt Brown,
Chief Scientific Officer, Genomics England

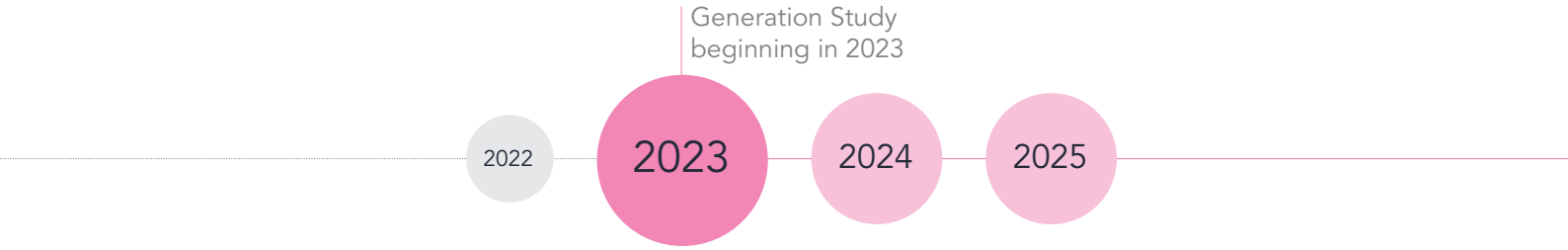
“Our ultimate goal in cancer is to maximise the benefits of genomic testing for patient clinical management by providing an up-to-date, accurate, comprehensive and easily interpretable view of clinically actionable biomarkers in a single WGS test with clinically relevant turnaround times. We are very lucky to be supported on this journey by the best minds from the academic and clinical research community in the UK and abroad. As a recent example, the knowledge hub that we have built around our unique collection of long-reads WGS for tumour clinical samples gave a critical boost to developing software for analysing Oxford Nanopore data with the high accuracy required for clinical testing.”

**Cancer**
Alona Sosinsky,
Scientific Director for Cancer, Genomics England

# Rare conditions

For rare conditions, we will increase the proportion of patients where a genetic diagnosis is identified and where a treatment is available by:

- assessing the utility of multi-omic approaches, including metabolomics, transcriptomics and proteomics, in combination with genomics to improve diagnostic accuracy
- improving the depth and accuracy of information associated with participants in the NGRL by increased linkage to clinical datasets as well as to 'omics and multimodal data housed in the NGRL, with analytical capabilities to make optimal use of that data
- continuing to adopt open community standards to facilitate data discoverability, interoperability and analysis, for example by using the Observational Medical Outcomes Partnership common data model
- working with academic and life sciences industry partners to develop improved diagnostic algorithms, including AI approaches, to better characterise variants of uncertain significance
- assessing new sequencing approaches including long-read and epigenetic sequencing to determine their diagnostic performance relative to the current gold-standard, short-read WGS
- supporting therapeutic development for rare conditions from target identification to supporting clinical trial programmes through case identification and synthetic controls
- improving genetic understanding of diverse ancestry patients and populations to improve diagnostic rates in these groups as well as to improve genetic interpretation in all populations
- delivering the Generation Study, which involves:[6]
    - evaluating the utility and feasibility of screening newborns for a larger number of childhood-onset rare genetic conditions in the NHS using WGS
    - understanding how babies' genomic data could be used for discovery research, focusing on developing new treatments and diagnostics for NHS patients
    - exploring the potential risks, benefits and broader implications of storing a person's genome over their lifetime




Aiming to find the

**9**

children born each day in the UK with a rare genetic condition – where early intervention is crucial

Expecting

**1,000**

positive results during the study

To find these children, we'll analyse

**100,000**

newborn genomes for a specific set of childhood-onset, actionable conditions

# Enabling research in areas beyond our focus

There are many opportunities throughout the patient journey where genomics research can fuel clinical transformation and help to achieve a positive patient outcome. We will maintain a strong focus on driving and enabling research in diagnosis and therapy development, particularly for rare conditions and cancer, which is where we are best placed to deliver benefits for research participants and patients. There are several other research approaches or areas that we may not lead on but will continue to support and enable through the data we make accessible to the research community:

### Common disease research

The growth of the NGRL, particularly with the addition of data from newborn WGS as part of the Generation Study, will enable researchers to conduct population-level genomics research into the risk factors and changes that lead to disease. Nonetheless, the data generation, acquisition, tooling and research we actively engage in and primarily support will remain on cancer and rare conditions.

### Functional genomics

We will enable this work by making genomic, multi-omic, multimodal and other research and clinical data and tooling available to clinical, academic and biopharma researchers. However, we will not directly engage in functional genomics studies or efforts to investigate the functional impacts of genomic changes.

### Clinical trials

Over the past decade, there has been a huge increase in therapy development for rare conditions and molecularly targeted cancer therapies. We are strategically well positioned to support both areas and enable clinical trials in them. By doing so, we advance therapies for major disease areas and increase the value of the genomic profiling we perform. This activity focuses on better linking patients with relevant trials – we are not involved in running clinical trials.

### Whole exome sequencing and gene panel data

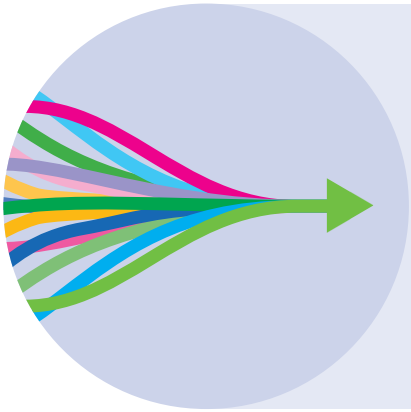
While we acknowledge that there is immense value in whole exome and gene panel sequencing data, in both clinical and research contexts, we are taking the long-term view that we will continue to focus primarily on WGS. By capturing an individual's entire genetic makeup, WGS offers the opportunity for a more holistic approach that will give deeper insights into complex genetic conditions. This is particularly relevant for conditions influenced by non-coding regions or structural variants, which may be missed by targeting whole exome or gene panels. Genomic knowledge is continuously advancing, and WGS offers the greatest potential for reanalysis and reinterpretation of data as new discoveries emerge. This aligns with our commitment to driving continuous innovation in genomic medicine through enabling diagnostic discovery and genomic research.

# Delivering on our vision

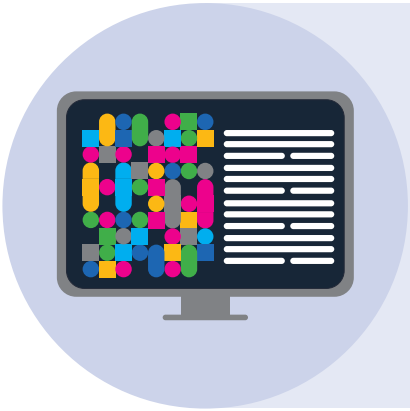
## Key aims

As an organisation, there are three key aims that we want to achieve together with the wider research community to realise this Research Vision:

1. **Improving our core research performance** by expanding the capabilities of our Research Environment and our data offer to researchers, by implementing changes to more actively enable collaboration and drive new research, and by supporting patient-participant involvement throughout the research process.



2. **Refreshing our governance** to ensure our processes and committees are set up so they can advise on, steer and govern the research activities we support or are involved in effectively.



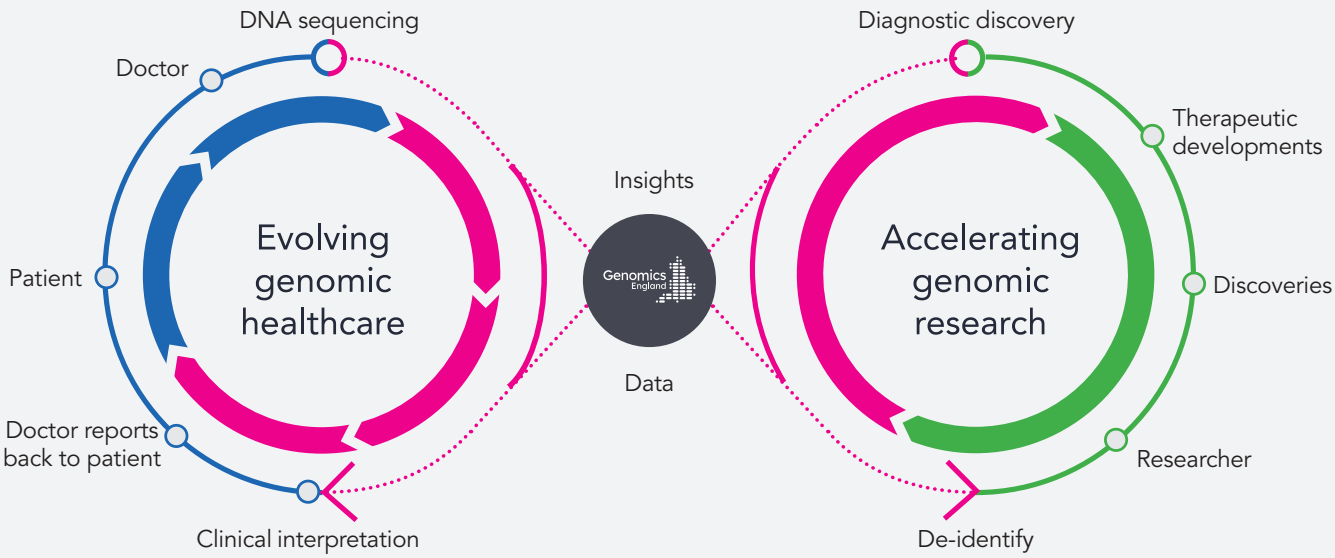
3. **Growing our impact** by expanding our focus beyond WGS to other technologies and data modalities (e.g. transcriptomics, proteomics, metabolomics, epigenetics, pathology and imaging data), as well as sharing our evolving understanding of these technologies, clarifying and building on our own research capabilities, increasing our international reach and strengthening partnerships across industry and academia.

By focusing on these, we will build a robust foundation to support, enable and drive research across rare conditions and cancer through to common diseases, fuelling the infinity loop of translational research and clinical impact.

# The infinity loop:
## connecting healthcare and research in genomics

**Evolving genomic healthcare**
We support the NHS to offer WGS to patients who might have genetic conditions and support Genomic Laboratory Hubs (GLHs) in the interpretation of results before they go back to a patient's clinician.

**Accelerating genomic research**
Many patients and their families opt to share their genomic and health data for research. We work together with the NHS to securely hold their data in the NGRL and facilitate secure access to it for approved researchers. They use the data to make scientific breakthroughs in rare conditions and cancer.



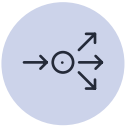

Healthcare data feeds into research, and research generates new insights that might improve diagnoses and treatments for NHS patients

# 1. Improving our core research performance

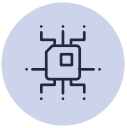
The NGRL is the main research repository to bring together genomics and related data (clinical, imaging and molecular) generated in partnership with the NHS. It is best positioned to maximise utility of this data to accelerate research and fuel clinical transformation. With the cost of generating and analysing WGS data for research purposes likely to fall significantly in the near future, as new technologies in both short- and long-read sequencing mature, improve in performance and become competitive commercially, new opportunities for applying WGS at scale will become feasible.

Collaborative partnerships are required to maximise these opportunities. Our goals are greatly facilitated by close engagement with the GLHs and the academic community, in both diagnostic discovery and research. Furthermore, patients and participants play a valuable and important role in guiding the design and performance of research programmes. We will seek new ways to improve their involvement throughout the research process, from design through to implementation.

To improve our core research performance, we will:



Improve the ease of use, support for and capabilities of the Research Environment



Increase the depth and quality of the data we make available to the research community

Improve our engagement with and support for our academic partner community, enabling more cross-cutting research and collaboration

Increase our ability to support research directly involving patients and participants

Improve patient-participant involvement across the research process

We will deliver on these commitments by:

- increasing the diversity of 'omics and multimodal data housed in the NGRL
- expanding the capabilities of the Research Environment over time to allow for the increasingly varied data the NGRL houses (eg RNAseq, proteomic, metabolomic and imaging data) and the analytical approaches our research partners are applying
- optimising the environment and associated support and resources to suit users with a broad range of backgrounds and analytic and computational expertise
- maximising collaboration, outputs and impact by more actively facilitating engagement with and among our research community to drive research and methodology projects relevant to our strategy and impact initiatives
- creating opportunities for individuals from different research areas to engage with one another, as well as facilitating connections between researchers and relevant participant communities
- expanding the Participant Panel and opening involvement opportunities to patients whose genomes have been sequenced through the GMS and to those who are contributing to our research programmes beyond the 100,000 Genomes Project, such as through the Generation Study and the Diverse Data Initiative

# 2. Refreshing our governance

Our goal is to ensure our governance processes provide appropriate levels of oversight and control. This will ensure research use of the NGRL is consistent with our policies and principles as well as patients' and participants' consent and expectations, while not placing unnecessary barriers and hurdles to academic or industry research partners. In refreshing our governance, we will strengthen and streamline our research governance processes, including through the optimisation of committee structures, remit and membership, and by making sure our governance and advisory committees include membership relevant to all of our research activities.

"While DNA sequencing is central to modern approaches to diagnosing rare conditions and guiding personalised cancer medicine, additional information can be gained by bringing together different 'omics approaches and other diagnostic modalities, such as imaging and histopathology. We are developing a large database combining WGS with digital histopathology and imaging data, initially from our 100,000 Genomes Project participants with cancer. This will enable research and development of multimodal approaches to personalised cancer medicine. We are also building multi-omic datasets in rare conditions, combining WGS short- and long-read data with transcriptomic and proteomic data in the first instance. Layering further data modalities, such as metabolomic and methylation data, will follow, providing additional opportunities for novel research into disease mechanisms and pathways.

Our ultimate goal is to build a deep dataset of participant clinical data together with comprehensive multi-omic profiling, with the primary goal of enabling research into the relative values of different 'omics approaches individually or in combination to increase diagnostic rates for our participants with rare conditions. These data will be invaluable for research into disease mechanisms, therapeutic target identification and functional genomics research across a wide range of human diseases."

**Multi-omics programme**
Greg Elgar,
Director of Sequencing, Genomics England

"It is increasingly accepted that cancer needs to be understood from both a molecular (genomic) and a spatial (histological and radiological) perspective, because both the molecular and morphological features of a tumour and the surrounding micro-environment have an impact on diagnosis, prognosis and response to specific therapeutic treatments.

As an example, we know that PD1 expression on the tumour cell surface and the presence of lymphocytes in the tumour micro-environment are both required for a successful response to immune checkpoint inhibitors. Insights such as these have tremendous research value, potentially identifying new druggable protein targets or new mechanisms of action, and they can also help to better stratify patients for clinical trial inclusion.

Current research insights are generated by isolated existing and mature technologies such as sequencing and digitisation of histology slides. Our vision is to combine analyses of these distinct data types to enable a greater understanding of disease and deliver better predicative models of diagnosis, prognosis and response to treatment.

The multimodal programme has an expanding digital image library of about 100,000 images, and we are developing open-sourced machine learning (ML) tools and pipelines with support from industry experts such as insitro."

**Multimodal programme**
Dr Prabhu Arumugam,
Director of Clinical Data and Imaging, Caldicott Guardian, Genomics England

# 3. Growing our impact

We believe that genomics, particularly WGS, will contribute more when combined with other approaches to both translational and basic research, and that this multimodal approach will move into clinical practice. Better enablement of genomic clinical services is already a key driver of internal Genomics England development programmes and joint iniatives with the NHS GMS, including bioinformatics pipeline development and genomic interpretation services, and we envision this will also deliver valuable research contributions across a diverse range of fields. We can greatly increase the use of and impact of our research assets, not least through international growth, and the contributions of research involving the NGRL can be maximised by increasing the number and diversity of researchers engaging with us.

To grow our impact, we will:

Enable multi-omic and multimodal research approaches

Better engage with non-European ancestry populations to increase the breadth of the population we partner with and support

Without diminishing our focus on rare conditions and cancer, grow our support of genomics research and clinical applications in common diseases in specific instances. For example, in pharmacogenomics, we are already laying the foundations for future research and clinical application in partnership with the UK Medicines and Healthcare products Regulatory Agency (MHRA)[7]

Build on existing programmes and partnerships to better enable AI research across basic science and translational applications including decision support for genomic medicine. In our collaboration with insitro, we are combining their ML–powered embedding search capabilities with our multimodal phenotypic and genetic research database[8]

Work with our partners to promote the advancement of therapy in both rare conditions and cancer from target identification and validation, through to encouraging and supporting clinical trials

Encourage and enable a greater number of researchers, including international researchers, to work with us and NGRL data

Deepen our engagement with life sciences industry partners to facilitate more academia-industry research collaboration

We will deliver on these commitments by:

- introducing 'omics datasets to complement the WGS data in the NGRL, increasing the number and range of digital histopathology datasets ingested, trialling and potentially adding radiology datasets to the NGRL, and adapting our databases and analytical platforms to better enable multimodal analysis
- broadening the research we enable, the research performed in active partnership with other organisations or with researchers where we are a collaborator, and research where we act as principal investigator. The latter will most likely occur when there are not yet suitable external partners to contribute significantly to areas where we see a critical research need. Identifying these areas will involve internal dialogue, discussion with academic and life sciences industry partners, engagement with the wider research and clinical communities, advice from our Scientific Advisory Committee, development with the NHS and GLHs, and the involvement of patients and participants
- increasing ease of access to and use of our Research Environment by improving the discoverability of datasets available for research, making it as straightforward as possible to become an approved user, providing user-friendly tooling, and continuously striving to improve the user experience
- working with our academic community and life sciences industry partners to foster biotech-academic engagement and collaboration, driving translational research programmes relevant to our core mission

# Evaluating our work

The potential for delivering significant healthcare benefits is enormous in the field of genomics. In pursuit of our research vision, we will continuously track progress, identify areas that need improvement and evaluate the overall impact of our research efforts to ensure they benefit the wider research community and research participants as well as leading to diagnostic and therapeutic advances. There are several ways in which we will measure progress toward these goals.

An important indicator will be the growth of our research community and establishment of partnerships with academic and life sciences industry partners. Collaboration with other organisations can facilitate the sharing of data, knowledge and resources, leading to more efficient and impactful research. In delivering on our research vision, we will not duplicate the activities of other public research agencies and organisations or compete with our research partners. Instead, we will strive to work collaboratively to enable research to achieve our collective goals. Patient and participant involvement in these initiatives remains a key aim, particularly where we play a convening role to further the development of new research. We will assess and evaluate the impact of our involvement approaches to ensure cohesive, useful collaboration between patient and participant representatives and researchers. Together these partnerships will increase the visibility and dissemination of the research that is being conducted within our Research Environment, ultimately leading to broader impacts.

Over **1,500** potential new diagnoses based on research on NGRL data were fed back to NHS clinical teams in 2022

Making the Research Environment more user friendly, streamlining processes for researchers to access and analyse data within it, and continuing to add large, linked datasets of considerable research value will lead to the number of researchers from academia and industry accessing the data held in the NGRL to continuously increase. This growth in our research community will result in a greater number of research projects being conducted in our Research Environment. We anticipate that over 1,000 research projects will be recorded in our Research Registry in 2024, and have introduced a new process to track research outputs and the impact these have.[9]

The ultimate goal of genomics research is to improve patient outcomes. Therefore, the success of our research operation can be measured by its ability to influence clinical practice and improve healthcare delivery. In partnership with the NHS GLHs we have established pathways to enable researchers to communicate discoveries that may be relevant to patient care back to the NHS, as well as routes to facilitating collaboration between researchers and clinicians to accelerate research efforts and their translation into practice. Over 1,500 potential new diagnoses based on research on NGRL data were fed back to NHS clinical teams in 2022. The successful delivery of this research vision will see the number of diagnostic discoveries and clinical-academic collaborations increase further year on year.

Alongside this direct route to patient impact, our close collaboration with policymakers and partnership with the NHS allow us to work together effectively to ensure that research on health data in the NGRL is translated into updates to our PanelApp and to the NHS GMS National Genomic Test Directory.[10] This will result in real-world benefits for patients.

We will continue to:

- Grow our community, particularly the number of researchers from around the world, with access to the NGRL via the Research Environment
- Advance the field of genomics and genomic healthcare through enabling a greater number and more diverse research projects to be conducted within our Research Environment, which can be tracked through the Research Registry and our annual Project Audit
- Deliver on dynamic partnerships, by being the global partner of choice for patients, clinicians and researchers, to work together to drive development of precision medicine approaches for rare conditions and cancer
- Work in partnership with the NHS to develop the route to feeding back potentially clinically relevant research findings so these can be further interpreted and clinically validated

# Meeting long-standing and emerging challenges

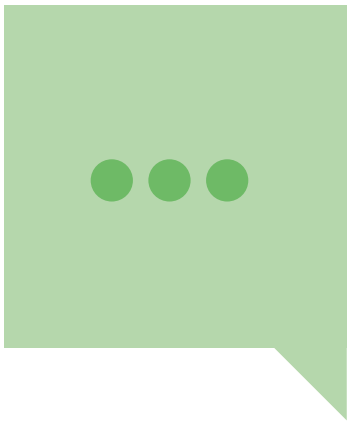
We have been at the forefront of genomic research, contributing significantly to the understanding of genetic variation and its role in health and disease. However, as the field of genomics continues to evolve, new challenges are emerging that require innovative solutions.

## Ethical challenges:
### continuously strive to set the gold standard for ethical research operations

Genomics research holds enormous promise for improving healthcare outcomes and advancing scientific understanding while also raising complex ethical, social and legal issues. Main concerns regarding genomics research include the privacy and security of personal genomic data. Patients and the public are rightly concerned about the potential misuse of or unauthorised access to their genomic information, which could lead to discrimination, stigmatisation or other harmful outcomes.

To address these concerns, we have developed comprehensive policies and procedures for data security, confidentiality and privacy, including de-identification of personal data for analysis within our secure Research Environment, stipulating the range of acceptable uses of data we hold and seeking to meet the expectations of participants who have consented to us collecting their genomic and health data. We will continue to work with others to develop policies and procedures for ensuring the protection of patient privacy and data security, while also enabling responsible and ethical research to benefit patients and the healthcare system.

In addition to ensuring we comply with research regulatory requirements, we draw on ethics advice and expertise to navigate new and emerging issues. We actively engage with patients, healthcare professionals and the public through a variety of communication and engagement channels to build awareness and understanding of the benefits and risks of genomics research. This includes bringing participant representatives onto key committees including the independent Access Review Committee and Ethics Advisory Committee.

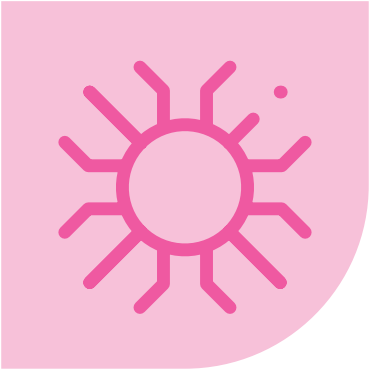


Another issue is the lack of diversity and representation in genomic research. As research has been predominantly conducted with people of European ancestry, the findings may not be generalisable to other populations with different genetic and environmental backgrounds. To address the dominance of European ancestry in genomic research, we are recruiting more diverse and representative patient cohorts, and engaging with under-represented communities to increase awareness and participation in genomic research through our Diverse Data Initiative.[11]

# Safe and impactful use of artificial intelligence:
## monitoring this rapidly developing and disruptive field to realise the transformative benefits it could deliver

The emergence of AI and ML presents both opportunities and challenges for the field of genomics. On the one hand, they hold significant potential for analysing and interpreting large volumes of genomic data, identifying patterns and predicting disease risks and treatment outcomes. This could lead to new insights and breakthroughs in genomic medicine, particularly where algorithms can be trained to identify patterns and associations that are not readily apparent to human researchers or clinicians. AI and ML approaches will also enable faster and more accurate diagnoses in many clinical indications, as well as help with identifying potential drug targets, predicting drug responses and achieving greater personalisation of treatments among many other advances.

However, AI and ML also pose significant challenges for genomics research. For example, the development of accurate ML models requires access to large volumes of high-quality data, which may be difficult to obtain in the case of rare or little-studied diseases. Additionally, AI and ML models can introduce biases if trained on data that is not representative of the broader population, leading to inaccurate predictions and potentially harmful outcomes.



We are actively exploring the use of AI and ML in our bioinformatic pipelines and research programmes, and are developing strategies to ensure the responsible and ethical use of these technologies. This includes working with participants, clinicians and researchers to collect diverse and representative genomic and clinical data, investing in robust data quality and privacy measures, and establishing best practices for the development and validation of AI and ML models in genomics research. The issues that AI and ML present in genomics research and healthcare require a collaborative and interdisciplinary approach, involving experts in genomics, data science and ethics. We believe that, through close collaboration and staying abreast of advances in the field, we can successfully leverage the potential of AI and ML to drive innovation and improved patient outcomes.

# Data storage and computational power:
## keeping up with the demands of large-scale interoperable datasets and complex analyses

The scale of genomic data is already vast, and genomic research and healthcare are generating large amounts of new data every day. We employ high-performance computing clusters, cloud-based storage solutions and advanced networking capabilities to enable us to store and process genomic and clinical data securely and efficiently. However, as the volume of data continues to grow, so does the challenge of managing and analysing it effectively. To address this, we are exploring new strategies to optimise our data storage and computing resources, such as using ML algorithms to identify patterns in data and enable faster, more accurate analysis. We are leveraging cloud computing to provide a scalable, more sustainable and cost-effective solution for large-scale and complex analysis. Our previously separate on-site and cloud environments have been merged to allow researchers using our Research Environment much greater flexibility in the types of analysis they can conduct on data in the NGRL.

With increasing data storage and energy costs, ensuring that access to the genomic and clinical data we hold remains free is another facet of this challenge. The cost of sequencing, curating, storing and analysing genomic data is significant, and we rely almost entirely on funding from public sources to make this possible at scale. At the same time, we recognise the importance of making data freely available to researchers worldwide, enabling collaboration and accelerating scientific progress. We want to continue to enable data sharing and collaboration across organisational and disciplinary boundaries, making it easier for researchers to build on each other's work and accelerate scientific progress. This requires ongoing investment in technology infrastructure and innovative approaches to data management as well as the development of sustainable models of data access and analysis.

"We apply ML to support both our healthcare and research missions in partnership with industry and academia. For evolving genomic healthcare, we bring ML to the biopipelines with the goal of improving and accelerating bioinformatic analyses. For accelerating genomic research, our main priority is developing multimodal ML models and use cases to support research, investigating ways to better characterise and treat cancers using integrated genomics, imaging and other clinical data.

ML also fuels our infinity loop (page 13) by tapping into our expertise across domains, such as analysing natural language to enhance data privacy and extract relevant information from clinical reports and scientific articles. We employ a combination of classic and cutting-edge deep learning techniques. Our ML engineers also bring a production-driven mindset and skills that are crucial for deploying trustworthy AI applications.

Overall, ML plays pivotal roles at Genomics England for unlocking insights from biomedical data that will ultimately enhance patient benefits."

**Machine learning**
Francisco Azuaje,
Director of Bioinformatics, Genomics England

"The Cohorts programme provides a route to enriching the Genomics England dataset for the research community across a range of cancers, rare conditions and population cohorts. While this asset will see a continued increase in genomic and clinical data arising from the NHS GMS (for those participants consenting to research), the clinical indications are limited to those commissioned.

To enable fundamental research in other disease areas, we partner with academic and life sciences industry collaborators through our Cohorts programme. We deliver a trilogy of components to generate data and fuel research, namely WGS, bioinformatics pipeline processing and access to the data within the NGRL. We work with our external cohort partners to support all steps of the cohort journey, from participant consent through to sample and data onboarding and, ultimately, data readiness.

We are delivering diverse cohorts, both disease focused and those expanding from newer areas of interest such as pharmacogenomics, as evidenced by our recently announced partnership with the MHRA Yellow Card biobank."

**Cohorts programme**
Shahla Salehi,
Service Owner – Scientific Research Delivery, Genomics England

“To date, genomics research has not equitably served all people, with substantial global populations excluded from research datasets and the research community. To fulfil our mission of bringing the benefits of genomic medicine to everyone, our datasets must accurately represent the diverse cancer and rare disease populations, as well as address common, yet previously overlooked conditions like maternal health complications and sickle cell disorder. By combining efforts to create more diverse datasets, implementing ethical research practices, and harnessing cutting-edge methods such as AI to work in service of equity challenges, we have a significant opportunity to comprehensively enhance diversity and equity in genomic medicine and research.”

**Diverse Data Initiative**
Dr Maxine Mackintosh,
Programme Lead - Diversity, Genomics England

# What does the future hold?

The increased capabilities of genomics point to an extremely promising future for how the field might advance healthcare and medical research. We envision a future in which WGS takes centre stage not just within clinical confines but also more broadly in the general community.

We see exciting potential benefits arising from WGS being available across a broader range of age and ancestry groups. We envision a 'genome for life' – a powerful asset guiding personalised medical journeys – being available for all. This compass could finely calibrate pharmacogenomic approaches, curbing the likelihood of adverse drug reactions and ensuring correct treatment dosages every time. Polygenic risk scores stand ready to unveil a new era, predicting disease susceptibilities and reshaping the medical landscape towards proactive early diagnoses and prevention.

We foresee advances in analytic approaches, embracing AI and advanced technologies for genome and epigenome analysis, greatly improving our ability to extract diagnoses and personalised disease management strategies from genetic data.

We envision genomic medical services in rare conditions and cancer being linked up with clinical trial networks so that people can be informed about potential new therapies from which they may benefit.

By advancing research together we can create this future where the benefits of genomic healthcare are available to everyone, regardless of their background.

# Appendix


1. **The Research Environment**
2023
www.genomicsengland.co.uk/research/research-environment

2. **Genome UK: the future of healthcare**
GOV.UK. 2020
www.gov.uk/government/publications/genome-uk-the-future-of-healthcare/genome-uk-the-future-of-healthcare

3. **Life Sciences Vision**
GOV.UK. 2021
www.gov.uk/government/publications/life-sciences-vision

4. **Data saves lives: reshaping health and social care with data**
GOV.UK. 2022
www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data

5. **Accelerating genomic medicine in the NHS**
NHS England. 2022
www.england.nhs.uk/publication/accelerating-genomic-medicine-in-the-nhs/

6. **Newborn Genomes Programme**
2023
www.genomicsengland.co.uk/initiatives/newborns

7. **MHRA and Genomics England to launch pioneering resource to better understand how genetic makeup influences the safety of medicines**
GOV.UK. 2023
www.gov.uk/government/news/mhra-and-genomics-england-to-launch-pioneering-resource-to-better-understand-how-genetic-makeup-influences-the-safety-of-medicines

8. **Insitro and Genomics England announce partnership to provide multimodal search capabilities and derivation of novel Insights**
2022
www.genomicsengland.co.uk/news/insitro-and-genomics-england-announce-partnership

9. **Research Registry**
2023
www.genomicsengland.co.uk/research/members/research-registry

10. **Genomics England PanelApp**
2023
https://panelapp.genomicsengland.co.uk

11. **Diverse Data Initiative**
2023
www.genomicsengland.co.uk/initiatives/diverse-data


Genomics
England