

Joining GeCIP and setting up data access

Contents

- [GeCIP eligibility](#)
 - [GMC members of staff](#)
 - [Patient and participant involvement representatives](#)
- [How to join a GeCIP domain](#)
- [GeCIP expectations](#)
- [Domain and institutional approval](#)
 - [Domain Access Review Committee approval](#)
 - [Participation Agreement](#)
- [Account issue](#)
 - [Information governance training](#)
 - [Signing GeCIP Rules](#)
 - [Accessing the Research Environment](#)
 - [Accessing clinical data in LabKey](#)
- [Accessing your home drive](#)
- [Principles of data access](#)
 - [Controlled and validated access to the dataset](#)
 - [Patient consent model and de-identification](#)
 - [Data security](#)
 - [Intellectual property](#)
 - [Publication of results](#)

GeCIP eligibility

Most academic researchers and clinical staff from across the world will be eligible to join the Genomics England Clinical Interpretation Partnership (GeCIP). Specifically, membership of GeCIP is open to any individual, student or member of staff who is affiliated with:

- UK academic research institutions (eg universities, research institutions etc)
- NHS trusts or authorities
- UK and foreign charitable organisations directly related to the focus of the 100,000 Genomes Project
- foreign universities and research institutions that carry out significant research activity
- UK and foreign governmental departments that carry out significant research activity (eg Medical Research Council, National Institutes of Health, Public Health England)
- foreign healthcare organisations (private or public) that undertake significant research activity.

Membership is not open to those who are employed by:

- private UK healthcare institutions
- commercial companies.

GMC members of staff

Members of staff from NHS Genomics Medicine Centres (GMCs) that have assisted with participant recruitment to the 100,000 Genomes Project are entitled to GeCIP membership, but are still required to complete the [GeCIP membership application form](#) ⚠️ so that they can be added to the membership of their chosen domain.

Patient and participant involvement representatives

The above restrictions on GeCIP eligibility may initially appear to exclude some patient and participant involvement (PPI) representatives from joining GeCIP. However, provided the individual concerned is not seeking to access the data directly within the Research Environment, they may request to opt-out of data access and remain a member of GeCIP.

How to join a GeCIP domain

There are details of [how to apply to join GeCIP](#) ⚠️ on the Genomics England website. In short, after filling in the online form, the applicant's details will be sent to the domain leads of the domains they selected in their application. The domain lead(s) will consider the applicant's skills and compatibility with the domain's research interests and decide whether to accept or reject the application. GeCIP has been founded with a 'team-science' mindset and the domain leads are encouraged to bear this in mind. The applicant will be notified of the domain's decision by email.

If you would like to join another domain or subdomain, please contact the domain or subdomain lead(s) listed on our [GeCIP domains](#) page to discuss your interest and how you may effectively contribute to the domain. Please note that membership of domains does not govern members' access to data – **all GeCIP members can see all de-identified 100,000 Genomes Project data within the Research Environment**. GeCIP members should only consider joining a new domain if their intended research activities fit within its scope.

GeCIP expectations

As a member of a GeCIP domain you will be given access to the 100,000 Genomes Project dataset. However, there are expectations of members and principles for research within the GeCIP research community that go beyond data access and analysis. These include:

- contribution to the knowledge base for genome interpretation (for example via [PanelApp](#))
- active and effective collaboration within your GeCIP domain and with other GeCIP domains
- expert assistance in genome interpretation to improve diagnostic yield for NHS GMCs
- contribution to training within your domain where needed
- contribution to the ongoing management and administration of the GeCIP domain where requested
- responsible and safe access to participant data
- submission to a Genomics England [Intellectual Property policy](#)
- other activities in line with the provisions of the [GeCIP Rules](#).

GeCIP researchers should note that their proposed research plans may overlap with the research plans of others both within their domain and, potentially, with other domains. Where such overlap exists, we encourage collaboration over competition. New domain members are asked to collaborate with those already undertaking the analysis.

Domain and institutional approval

Once a member of a GeCIP domain, there are a number of steps through which the researcher (and the domain itself) must pass before being granted data access.

Domain Access Review Committee approval

Prior to being considered for data access, each GeCIP domain must submit a detailed research plan to the [Access Review Committee](#) (ARC) who will consider whether the research plans of the domain are in line with the stated acceptable uses of the Genomics England dataset. The acceptable uses are outlined in the Research Ethics Committee (REC)-approved [100,000 Genomes Project Protocol](#) and the [Data Access and Acceptable Uses Policy](#). The ARC will decide to approve, decline or amend requests for access to data. The domain leads, or a deputy, will be asked to call in at the appropriate time to answer any questions or queries, and external expertise may be called upon to assist the decision-making process, where appropriate.

Participation Agreement

Each institution that is represented within GeCIP is required to sign a contract, the [Participation Agreement](#), with Genomics England. The provisions of the agreement are manifold, but in particular it outlines the various obligations of both the institution and Genomics England with regard to the data access granted to that institution's researchers. It also deals with the ownership of intellectual property (IP) resulting from research on the 100,000 Genomes Project, citing the project's [IP Policy](#). To promote a fair, open and co-operative environment, all GeCIP members are expected to work within the agreed terms of the Participation Agreement signed by their host institution and abide by the terms outlined in the [GeCIP Rules](#). Detailed information on these and other documentation related to GeCIP guidance and policies can be found on the [GeCIP Documents and Resources](#) page. A list of institutions and whether they have signed the Participation Agreement can be found [here](#).

During the signing process, each institution will provide a verified list of GeCIP members from their institution to Genomics England. The complete list of users and access requirements is then processed by the Genomics England Service Desk. You must be verified by your institution in order to gain access to the project dataset.

Account issue

When it is time to give you access to the data, a Genomics England account will be issued to you using the institutional email address that we hold for you. You will receive two emails:

1. An email with your username and a link to complete your information governance training
2. An email with a link to set your password.

The same username and password will be needed to log in to applications, eg LabKey, within the Research Environment individually. The safety of your password is of utmost importance.

- **Try to be a little unusual when adding special characters.** Replacing an s with \$ and a ! on the end are very common, and is something a person will try if they think they might know your password. When was the last time you put a # into a password? Just adding one random character can make your password much more secure.
- **Never write it down**, except in a secure password manager. If you write down your password, it defeats the purpose of having one entirely. If you need clues as to which password you have used to each system, keep a log of **clues**, but make sure they do not contain anything to do with the password (For example, if my password is '@range\$86' – my clue could be 'Fruit' or 'Vitamin C') - **Spend time** to make a good quality password and clue.
- **Never share it** – if you need to for some reason share a password protected document or system with someone else, change the password to something specific to that task. **NEVER** share your own password.
- **Public or shared computer? Do not SAVE your password.** Be mindful when letting a system remember your password. You could potentially be giving other people full access without the need to guess anything!

Once you have your login credentials, there are a number of steps to be undertaken before you will get data access:

1. You will need to complete an Information Governance course.
2. You will need to sign the GeCIP Rules agreement.

3. You will need to wait for email confirmation that you have completed IG training and can now access the Research Environment. Please note, this confirmation can take up to 5 days.

Information governance training

To protect patient confidentiality and ensure that all researchers understand the sensitivity of the data and how to maintain its security, we ask that all individuals requesting access to the Genomics England Research Environment undertake a course on information governance. Data access will not be granted until a researcher has successfully passed information governance training.

You must first access the Genomics England Moodle training environment. There will be a training course entitled 'Information Governance and GeCIP Rules'. Select this course and work through the pages and final quiz.

Signing GeCIP Rules

Signing of the GeCIP Rules, an agreement to which GeCIP researchers must agree to in order to gain data access, will be done following IG training in the Genomics England Moodle environment.

Accessing the Research Environment

To access the Research Environment, you should open a browser and navigate to <https://re.extge.co.uk/ovd/>. You will then be provided with a login dialog box for which you should use the username and password that have been issued to you. If this does not give you access then it would be worth trying a different browser, eg either [Google Chrome](#) or [Mozilla Firefox](#). If you're still unable to get access, please submit an issue to the [Genomics England Service Desk](#) stating your username, the operating system you are using, the different browsers you have tried and any error messages that are shown.

Accessing clinical data in LabKey

To access the clinical data, you need to:

1. log into the Research Environment
2. open LabKey by double-clicking on the [LabKey icon](#) on the desktop
3. sign in using the same credentials you used to access the Research Environment by clicking 'sign-in' in the top-right corner of the browser window
4. at this point, you should see a folder called 'MAIN PROGRAMME', which contains each data release from the Genomics England main programme
5. if you are taken to an empty LabKey page and no data are visible, submit an issue to the [Genomics England Service Desk](#) entitled 'GeCIP researcher <username> requesting LabKey access' stating that you have successfully logged into LabKey within the Research Environment but do not have access to any data. You should state you are a GeCIP member and need access to the main programme data. You will be notified when this has been completed, and the next time you log into LabKey you should have access to the appropriate data.

Accessing your home drive

Your home drive contains links to shared files:

- genomes - gives you access to the Genomics England datacentre and all of the sequencing data.
- re_gecip - contains the shared folders for each of the GeCIP domains. Access to the individual GeCIP domain folders is restricted to members of the domain.

Principles of data access

The Research Environment is provided within a framework of underlying principles that define its nature and design. They are driven by the agreements under which the data has been made available and control how you can use the data.

Controlled and validated access to the dataset

As you will know from the process you have gone through to get access to the Research Environment, access is granted based on a set of rules:

- Researchers (individuals, students or members of staff) must be affiliated with an institution that meets the criteria for a research organisation and that has itself signed up to Genomics England's access agreement.
- The research project must be real, justified in terms of patient outcomes and be part of a GeCIP domain.
- The individual researcher has been subject to due-diligence processes by Genomics England.

It is through these requirements that we are able to maintain our commitment to patient anonymity and security, and to gain the highest possible value from the 100,000 Genomes Project dataset.

Patient consent model and de-identification

While Genomics England has a patient's data, identifiers such as NHS number and postcode are removed to protect their identity and to make it harder for anyone to attempt re-identification of clinical and genomic information for a particular individual.

The conditions of access to the Research Environment include a commitment to not attempt to re-identify a patient by reverse engineering a health record.

It is not assumed that a particular patient's data will remain part of the dataset; they are entitled to withdraw their consent for their data to be used at any time.

Data security

To ensure a high standard of data security and integrity, access to the Research Environment is only granted when you, the researcher, have undergone mandatory information governance training. This involves undertaking an online training session and completing an online test. To access the Research Environment, you must achieve a pass mark of at least 80%.

Researcher activity and behaviour within the Research Environment are monitored. This enables Genomics England to define and maintain the level of standards that will encourage more patients to want to be included in the project and to attract high-calibre researchers from around the world to participate.

Intellectual property

Membership of a GeCIP domain and access to the Research Environment is on the understanding that all outputs of the research and all IP are automatically wholly owned by Genomics England where the research has been conducted entirely within the GeCIP programme and when it does not use assets owned by third parties. In other circumstances, where non-GeCIP assets are involved, IP ownership will be negotiable.

It is important that all research outputs and IP be fully disclosed to Genomics England before publication.

There is a separate document, [Genomics England IP Policy](#), that you are encouraged to read as it includes sections on patents and the licensing of IP.

Publication of results

Genomics England encourages the publication of results - indeed, it is considered an essential aspect of the 100,000 Genomes Project that the results and discoveries resulting from your research are published promptly.

However, Genomics England has in place a 9-month moratorium on the publication of results to provide a degree of preferential permission to publish on participant data. This is in recognition of those domains and individuals that have assisted in recruiting to the project. In effect, the moratorium will prevent other GeCIP researchers from publishing research on that domain's corresponding data until the domain has had access to the data for at least 9 months. The moratorium begins when a meaningful number of whole genomes is available to researchers in a specific domain. More detail on the moratorium can be found in the [Genomics England Publication Moratorium](#) policy.

Given that the research will be based on work carried out in the Genomics England Research Environment, and within a GeCIP domain, all publications will be made 'on behalf of the 100,000 Genomes Project and Genomics England' as a banner heading. In general, issues relating to authorship (whom to include, the order of authors, etc) will follow the guidelines issued by the International Committee of Medical Journal Editors (ICMJE ). For a full description of how and when to publish your results, please refer to the [Genomics England Publication Policy](#). It includes detailed sections on authorship, patentability and review.