Reanalysis of genomic sequencing data for rare disease patients in the Genomic Medicine Service

Document name:	Reanalysis of genomic sequencing data for rare disease patients in the Genomic Medicine Service
Document number:	GU-LAB-002
Author	Sandi Deans, Rachael Mein
Version	2.0
Date:	September 2023
Review period	Annually

Document management

Revision history (draft versions)

Version	Date	Author	Summary of changes
0.1	11/06/2021	Sandi Deans	
0.2		Rachael Mein	Incorporated feedback from GLH Partnership board Inserted section on reanalysis of non-WGS data Included process for Diagnostic Discovery

Revision history (final version)

Version	Date	Author	Summary of changes
1.0	12.01.2023	Sandi Deans	First final publication after consultation of draft
2.0	28.09.2023	Rachael Mein	Updated 100K reanalysis to be performed by GEL Diagnostic Discovery analysis Updated 100K additional analysis to enable submission for testing via GMS Added appendix related to GEL DD analysis and its consequences on reanalysis policy

Reviewers

This document was reviewed by the following people:

Reviewer name	Title/responsibility	Date	Version
GLH Partnership board		July 2021	0.1

Approved by

This document must be approved by the following people:

Name	Title	Date	Version
Prof Dame Sue Hill	SRO for Genomics, NHS England and NHS Improvement	28.09.2023	2.0
Prof Sandi Deans	Deputy Director (Laboratory & Scientific), Genomics Unit, NHS England and NHS Improvement	28.09.2023	2.0

Document control

The controlled copy of this document is maintained by NHS England. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Contents

1 Executive summary 4

2 Introduction 5

3 Reanalysis/additional analysis for individuals referred for WGS through the GMS 6

3.1 Reanalysis requests 7

3.1.1 For cases originally sequenced via Genomics England 7
3.1.2 For cases originally sequenced by a GLH 8
3.2 Additional analysis requests 8

3.2.1 For cases originally sequenced via Genomics England 8
3.2.2 For cases originally sequenced by a GLH 9

4 Reanalysis/additional analysis for individuals who participated in the 100,000 Genomes project 10

5 Diagnostic discovery 12

6 Reanalysis of GLH derived Genomic data 15

6.1 Reanalysis following addition of “Green genes” to panels 15

6.2 Reanalysis of R21 exome sequencing data through the R14 rapid genome sequencing service 15

6.3 Reanalysis of R14 exome/genome sequencing data 15

6.4 Reanalysis requests where family members are considering having a child or a diagnosis could aid clinical management. 16

7 Roles and Responsibilities 17

8 Appendix 18

9 References 22

1 Executive summary

Patients with rare and inherited diseases, whom have no confirmed (or only partial) diagnosis to date and have been referred for Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) through the NHS Genomic Medicine Service (GMS), recruited to the 100,000 Genomes Project and/or identified through additional research projects, may benefit from reanalysis of existing genomic data (e.g. WGS or WES data) to identify previously undetected findings.

The advantage of reanalysis is that the patient's DNA sample does not need to be re-sequenced, however there is a responsibility for the NHS to validate the findings, when appropriate, and for a relevant clinical team to recontact the patient with an updated result. Reanalysis will not be performed in an *ad hoc* manner and will follow defined criteria and approval processes.

At present it is not possible to utilise WGS data generated during the 100,000 Genomes Project in analysis performed within the GMS WGS service. Therefore, if the clinical question has changed and/or different analysis parameters from those originally requested are required, patients should be re-tested via tests available on the National Genomic Test Directory.

Re-analysis of WGS data (processed via Genomics England) to make use of updated knowledge for the same clinical question is currently being performed via Genomics England's internal Diagnostic Discovery program. It is proposed that more automated ways to perform reanalysis will be explored by the Rare Disease Network of Excellence.

The governance on reanalysis of non-WGS data, or WGS data generated within a GLH, is the responsibility of the GLH who generated the raw data. However, the principals set out in this document, on when and why reanalysis could be considered appropriate, should be followed across the national network. Any reanalysis will be performed by the GLH who generated the genomic data, however, they may decide to share the data with another GLH to make use of alternative bioinformatic pipelines or analysis strategies.

Reanalysis will be performed as per the National Genomic Test Directory codes;

- **R387** Reanalysis of existing data – using additional gene panels, where new phenotype data is available, following bioinformatics improvements etc. - see sections on Reanalysis of; GMS Genomics England WGS data (section 3), GLH derived non-WGS/WGS data (section 6) and data from the 100,000 Genomes Project (section 4).
- **R447** Diagnostic Discovery[#] – when a putative diagnostic variant has been identified but validation or verification of its presence and/or variant classification and reporting in a diagnostic setting is appropriate - see Diagnostic Discovery (section 5).

[#] For recording of activity via patient level contract monitoring (PLCM) a TESTMETHOD should be recorded when confirmatory testing is performed and TESTREPORT should be used for variant classification and reporting. Where confirmatory testing is not required only TESTREPORT activity should be recorded.
R447 is a new code added to the Test Directory in October 2023. Prior to this R370 Validation Test code was used for this purpose.

2 Introduction

The NHS GMS is the national provision of comprehensive genomic testing for rare and inherited disease and cancer within NHS England and is informed by the National Genomic Testing Directory which sets out the tests that are nationally commissioned and funded.

This document defines the process, in the NHS GMS, of reanalysis of genomic sequencing data for patients with rare and inherited disease with no confirmed diagnosis, or only a partial explanation.

For the purpose of this document, the definition of '**reanalysis'** will henceforth refer to scenarios where the original clinical question is unchanged and reanalysis is requested to make use of bioinformatic improvements, gene panel updates or updates in genomic knowledge such as pathogenic variants published in ClinVar.

The alternative term '**additional analysis'** is used when a subsequent request is made following submission of a patient/family for WGS, to analyse additional gene panels or change clinical information that affects the analysis as it was incorrect, incomplete or has changed since the case was originally submitted. Where possible scenarios that require additional analysis due to incorrect/incomplete clinical data including not considering all known relevant gene panels from the beginning, should be avoided by checking the clinical information submitted is complete and accurate and that all possible relevant gene panels have been applied at the offset. This is to ensure efficient use of resources by reducing the amount of additional analysis requests necessary and to avoid delays in issuing the full test required for the patient/family.

However, updated knowledge can achieve additional diagnostic yield through applying it to existing WGS or WES genomic data. Much of this comes from new gene identification and a more comprehensive understanding of normal and pathogenic genomic variation over time, although it may also result from new clinical information (evolution of the patient's phenotype, or new non-genetic test results). Diagnostic yield may also be higher with substantial improvements to genome sequencing bioinformatic analysis, such as inclusion of the analysis of Copy Number Variants (CNVs) and Short Tandem Repeat Sequences (STRs), improvements in sensitivity of other elements of the pipeline including variant calling, detection of structural variants (SVs), deep intronic variants (Flanagan, 2013) (Ellingford, 2018), 5' and 3' regulatory variants (Weedon, 2014) (Evans, 2018) and mobile element insertions (Kvikstad, 2018). Therefore, it is acknowledged that some reanalysis and additional analysis is required particularly with a changing phenotype or significant increase in knowledge and improvement of analysis processes.

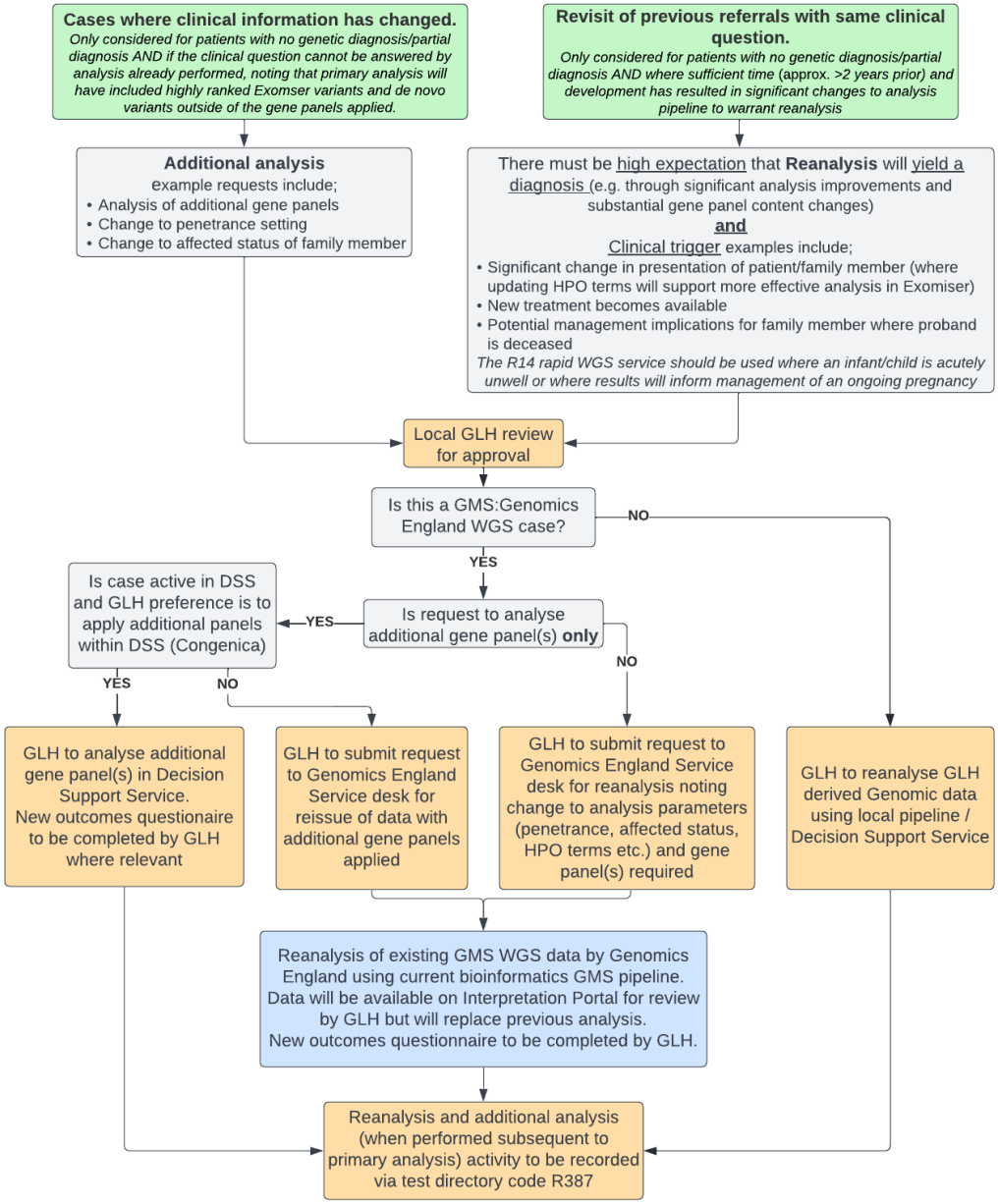
This document sets out the circumstances where reanalysis and additional analysis of existing data may be triggered by a clinician, on clinical grounds, or where a likely diagnostic variant is subsequently identified that warrants return to the NHS for formal review. This document sets out the potential sources of such variants and the actions to be taken.

3 Reanalysis/additional analysis for individuals referred for WGS through the GMS

Clinical reanalysis/additional analysis requests for rare disease patients must be made by the Clinician managing the patient/family and referred to the local GLH. Any request to reanalyse genomic data or perform additional analysis will be reviewed and, if appropriate, approved and processed by the GLH. The GLH shall have a process for approval of such requests which will include input from appropriate Clinicians and Clinical Scientists within the GLH.

For cases with data from the centralised WGS service, GLHs should **review any findings identified through the Diagnostic Discovery pathway before reanalysis or additional analysis is approved**.

Figure 1: Pathways for reanalysis/additional analysis for GMS referrals



3.1 Reanalysis requests

Efficient reanalysis requires a semi-automated approach which is yet to be established. Therefore, **GLHs are expected to perform very little, if any, reanalysis** of GMS centralised WGS data. The internal Genomics England Diagnostic Discovery analysis work will bridge this gap to look for variants in new genes added to panels or to address previous pipeline limitations (see appendix for more detail).

Reanalysis of WGS data should only be performed where there is an immediate clinical need and where reanalysis is likely to increase the chance of a diagnosis due to new gene associations, additional clinical information, improved bioinformatics etc. (see Figure 1). It must not be performed in an uncontrolled *ad hoc* manner nor when it is highly unlikely to generate additional results from the primary analysis.

Requests for WGS data reanalysis may be accepted as appropriate if the following criteria are met;

1) **There must be a high expectation that reanalysis will yield a diagnosis.** For example;
   - the change to gene panel content or bioinformatic improvements introduced since the last analysis was performed, is significant so that the probability of achieving a diagnosis is sufficient enough to justify reanalysis. It is not expected that reanalysis will be significantly productive within at least 2 years of the last analysis.

AND

2) **There must be a clinical trigger for the reanalysis**. For example;
   - There is a new significant change in the clinical presentation of the patient or family member (e.g. evolving phenotype) meaning application of additional GMS approved gene panel(s) would be relevant and updating the HPO terms provided will support more effective analysis in Exomiser;
   - New treatment becomes available so reanalysis could influence the patient's clinical management;
   - The proband is deceased and the family initiate further investigation through the appropriate clinician, due to potential management implications for themselves and/or the wider family.

3.1.1 For cases originally sequenced via Genomics England

All locally approved requests for clinical reanalysis of WGS data should be submitted to the Genomics England Service Desk. Cases will be analysed using the current Genomics England bioinformatics pipeline and will subsequently be available for reanalysis by the GLH via the Interpretation portal. At present, it is not possible to distinguish reanalysis cases from cases undergoing primary analysis in the interpretation portal as reanalysed cases are dispatched for interpretation via the same route as new cases. Upon reanalysis by Genomics England **the original data displayed and recorded in the Interpretation portal and decision support service will be replaced with the new data therefore previous decisions will not be visible** so any relevant information from the original analysis should be noted before the data is over-written so that it can be included in the reanalysis where required.

3.1.2 For cases originally sequenced by a GLH

Each request must meet the criteria above and should be submitted to the GLH who performed the sequencing using a specific request form (R387 Reanalysis request form). The sequencing GLH will agree the reanalysis strategy with the Clinician making the request. Cases will be analysed using the sequencing GLH's current bioinformatics pipeline. Clinically urgent reanalysis requests will be reported within 2 weeks of the reanalysis approach being agreed between the sequencing GLH and the requesting clinician. For further details see Section 6.

3.2 Additional analysis requests

This section refers to cases whereby additional analysis is required to answer the ***current*** clinical question. Examples of additional analysis requests include;

- Centralised WGS R98 originally requested but R27 subsequently requested
- R14 WGS 'gene agnostic' performed but panel test required as reduced penetrance is plausible in family.

3.2.1 For cases originally sequenced via Genomics England

Every effort should be made to avoid the need for additional analysis by providing high quality information at the point the patient/family is referred for centralised WGS and upon submission to National Genomic Informatics Service (NGIS), ensuring all relevant gene panels are requested.

Additional analysis requests as a consequence of incorrect clinical information that affects the analysis (e.g. penetrance or disease status for any family members such that tiering is based on incorrect segregation patterns) should be submitted by the GLH via a service desk request to Genomics England requesting reissue of the data under the corrected parameters.

Where additional gene panels are requested, it should first be considered if the analysis already performed is likely to have detected causative variants outside of the panel applied. Current analysis strategies already include looking for *de novo* variants in Trios and top Exomiser ranked variants (variants ranked by predicted deleterious impact in genes most often associated with HPO terms provided) in genes not on the panels applied. **The overlap in gene panels should also be considered to evaluate if the majority of genes on the new panel requested have already been analysed and therefore the additional analysis required may be of limited benefit and diverts resources away from those patients who have not yet had a primary analysis.**

If additional gene panel analysis, following the primary analysis, is still required, there are 2 options from which the GLH can decide how best to proceed;

1) If the case is still active in Congenica and the GLH is informed that additional gene panels are required, then these can be applied by the GLH in Congenica (see (Guidelines for Rare Disease Whole Genome Sequencing & Next Generation Sequencing Panel Interpretation & Reporting)). The benefit of analysis of additional gene panels within Congenica is that additional analysis can proceed immediately

with no intervention from Genomics England and any previous variant annotations within the audit tab are retained (this is not the situation if the case is reissued as the data has to be removed, reissued and re-dispatched to the DSS). However, it is not possible to add further HPO terms and update Exomiser within Congenica and the analysis of additional panels within Congenica is not captured within the interpretation portal therefore where appropriate the additional panels applied should be documented in the outcomes questionnaire.

2) If the case is no longer active in Congenica, the GLH has not started the primary analysis so its removal and redispatch to the DSS will not lose any variant analysis decision already made or the GLH deems a reissue is the most appropriate way to provide this additional analysis, then the GLH should contact Genomics England and request that the case is reissued with the additional gene panels applied. Please note that upon reissue by Genomics England **the original data displayed and recorded in the Interpretation portal and decision support service will be replaced with the new data therefore previously applied panels should be included in the additional analysis**, and any relevant information from the primary analysis should be re-entered on the new outcomes questionnaire It is not possible to change the primary clinical indication but additional panels relevant to other clinical indications can be applied.

The standard time from receipt of the request from the GLH to the re-processed data being released from Genomics England will be within 6 weeks. Each request will be treated under the standard pathway, unless they are clearly flagged as clinically urgent with the rationale for urgency outlined with the request, for example where the new information is being sought to inform on-going pregnancy management or transplant decisions. The time from receipt of urgent requests from the GLH to the re-issued data being released from Genomics England will be within 2 weeks.

When additional analysis is performed a new outcomes questionnaire should be completed even if there are no new findings from the additional analysis. Previous findings should be included to ensure the data retained includes both the initial review and the additional analysis review as only the latest version of this is displayed in the Clinical Variant Ark (CVA), although all previous data is available via audit logs upon request.

Additional analysis activity should be recorded for Patient Level Contract Monitoring via the test directory code R387: Reanalysis of existing data (see Figure 1).

3.2.2 For cases originally sequenced by a GLH

For additional analysis of genomic data generated by a GLH see section 6 for more details.

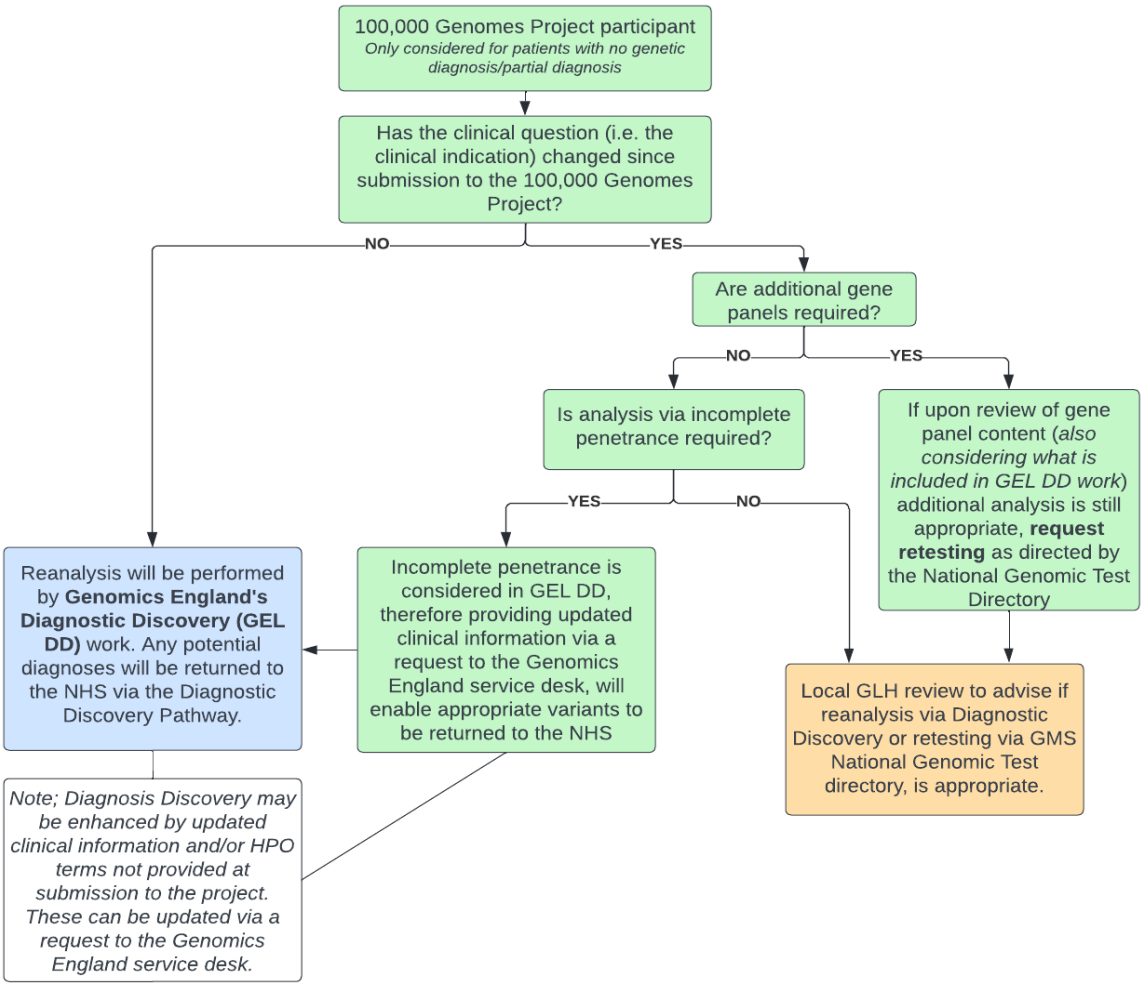
4 Reanalysis/additional analysis for individuals who participated in the 100,000 Genomes project

If participants have not had a finding result from the 100,000 Genomes Project and they have no clear reason for additional analysis (there is no new phenotypic information, no immediate clinical action, and the initial clinical parameters like; penetrance, pedigree and clinical data, were correct), then it is **strongly recommended that reanalysis for these individuals is deferred to the Genomics England internal Diagnostic Discovery analysis work**, which aims to streamline the use of updated pipelines and gene panels for reanalysis, making use of a cohort wide approach and focusing on areas where high throughput pipelines have limitations (see appendix 1 for more details). Diagnosis Discovery may be enhanced by updated clinical information and/or HPO terms not provided at submission to the project. Where appropriate these can be updated via a request to the Genomics England service desk.

Where additional analysis is required (e.g. different genes panels to those analysed during the project) it is acknowledged that this is challenging within the 100,000 Genomes Project Interpretation Portal due to the lack of decision support tools and not all 100,000 Genomes Project disorders have an equivalent or relevant gene panel to those within the NHS GMS. It is also not currently possible to move WGS data from the 100,000 Genomes Project into the GMS WGS system. Therefore, if the GLH decides that **additional analysis is required then these cases should be tested through the relevant test on the National Genomic Test directory** (including retesting via the GMS WGS pathway if clinically eligible). Additional analysis, by retesting via the GMS, should only be requested if the clinical question is different or extended from that submitted to the 100,000 Genome project and hence not covered by Genomics England's Diagnostic Discovery. The overlap in different gene panels should also be considered to evaluate if the majority of genes on the additional gene panel(s) requested have already been analysed and therefore the additional analysis required may be of limited benefit and diverts resources away from those patients who have not yet had a primary analysis.

Potential diagnoses are continually being identified by research activity within the National Genomic Research Library, in addition to Genomics England's Diagnostic Discovery analysis, and returned to the NHS GMS for review and reporting via the Diagnostic Discovery pathway (see section 5). Therefore, there may be times when the GLHs have been notified of a Diagnostic Discovery but the patient's clinician is not aware of it so **GLHs should check if there is a Diagnostic Discovery pending before initiating any additional testing**.

Figure 2: Pathways for reanalysis of 100,000 Genomes project participants data



Outlined below are examples where it is acceptable to test patients from the 100,000 Genomes Project, and their family members (where appropriate) through the relevant test from the GMS National Test Directory (including WGS if clinically eligible);

- **Patient submitted to 100,000 Genomes Project as a singleton but it is now possible to test as a Trio** (for clinical indications whereby trio analysis is the optimal family structure) and the primary analysis detected lots of potential candidate variants which trio analysis will help to classify.
- **There is an expectation that a new test would yield a new diagnosis**, e.g. the original analysis did not consider the full phenotype or the phenotype has substantially changed. The patient can be tested through the relevant test from the GMS National Test Directory (including WGS if clinically eligible).

If samples from the patient and their family members are submitted for re-sequencing via the GMS WGS pathway it should be documented, with relevant 100,000 Genomes Project reference, on the patient notes field within the National Genomic Informatics Service (NGIS) and in the outcomes questionnaire, that the patient/family were a participant in the 100,000 Genomes Project. This information will then be available within the Test Order Management System (TOMS) and CVA to inform users within the GMS that previous WGS analysis has been performed and flag any potential duplicate entries in CVA.

5 Diagnostic discovery

Diagnostic Discovery is a pathway developed by NHS England and Genomics England to ensure that potential diagnostic variants, discovered through investigations/research using the National Genomic Research Library (NGRL) and from internal Genomics England Diagnostic Discovery analysis work, are assessed and where relevant interpreted and reported, within the NHS GMS, in order to achieve new diagnoses in patients for whom there is WGS data from the 100,000 Genomes Project and the GMS WGS service.

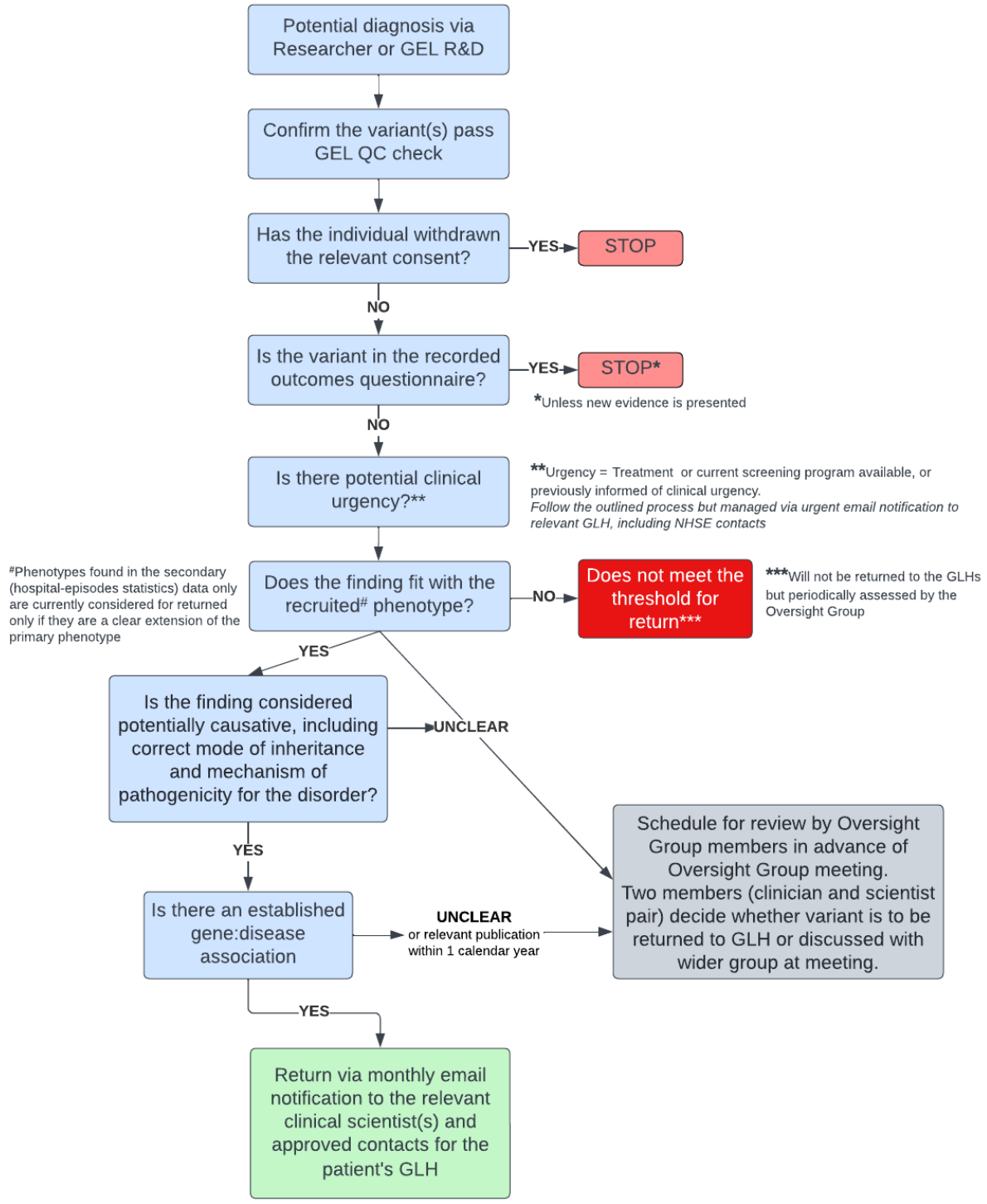
Sources identifying new genomic findings, and therefore potentially diagnostic variants, include:

1) Researchers: Including academic researchers from GeCIPs (Genomics England Clinical Interpretation Partnerships) and Discovery Forum commercial partners,
2) Internal Genomics England research and development (including bioinformatics) during validation, evidence generation, development and iteration of rare disease pipelines as well as the ongoing Diagnostic Discovery analysis work.

Potential diagnoses are pre-screened by Genomics England according to rules developed with and agreed by the NHS Diagnostic Discovery Oversight group. The Oversight group comprises of predominantly Rare Disease clinicians and clinical scientists nominated to represent each GLH. When required the Oversight group will seek additional specialist expertise to aid discussion of certain cases.

Figure 3 outlines the Diagnostic Discovery triage process. This involves categorisation of variants, by the Clinical Research Interface team at Genomics England, into those which meet the threshold for direct return to the GLH and those which require scrutiny via the Oversight Group. The latter comprise cases where there is uncertainty about the phenotypic relevance or validity of the variant(s) and where there are new gene/phenotype associations (association reported within the last 12 months) in addition to any cases considered likely to be incidental findings. Upon review the Oversight Group decide which of these variants should be returned to the GLH.

Figure 3 Diagnostic Discovery Process map



Pertinent Diagnostic Discovery variants, which are returned to the GLHs, should be classified using current ACMG (Richards, 2015), ACGS (Ellard, 2020) criteria and when required confirmed by an appropriate accredited test before reporting. Interpretation and Reporting guidance (Guidelines for Rare Disease Whole Genome Sequencing & Next Generation Sequencing Panel Interpretation & Reporting) should be followed regarding variant investigation, validation and orthogonal testing.

An outcomes questionnaire should also be completed within either the 100,000 Genome Project Interpretation portal or the GMS Interpretation portal, depending on how the patient's sample was submitted for WGS. This will ensure that patient/participant's variants are not fed back into the Diagnostic Discovery pathway again as it will be clear they are already known to the GLH and will also support the "Known Pathogenic Variant Prioritisation" (KPVP) feature within the GMS WGS pipeline.

For more information on the Diagnostic Discovery pathway see (NHS Diagnostic Discovery Guidance). For more information on the Genomics England Diagnostic Discovery analysis work see Appendix 1 of that guidance.

6 Reanalysis of GLH derived Genomic data

This section addresses the expectations of the GLHs in relation to reanalysis of any genomic data generated by a GLH for which updated analysis tools/gene panels could result in reaching a final diagnosis which previously was not achieved. This primarily refers to, but is not limited to, WGS, WES and large targeted capture Next Generation Sequencing (NGS) panels.

6.1 Reanalysis following addition of "Green genes" to panels

There is no expectation to retrospectively reanalyse new Green genes in all patients tested via a previous panel version. Similar to reanalysis of the centralised WGS data via Genomics England, clinical reanalysis requests must be made on a case by case basis by the Clinician managing the patient/family and should take into account the probability of achieving a diagnosis versus the resources required to support the additional analysis. GLHs should also assess such requests to ensure they are proportional to the diagnostic yield and technically feasible based on the data available. The GLH shall have a process for approval of such requests. The analysis of additional genes should be appropriate for the patient's phenotype and the patient must fulfil the current eligibility criteria for that clinical indication.

The same expectation applies for tests delivered via a targeted capture approach where genomic data is not available on genes added to new versions of panels. Retesting should only be performed if there is high probability of achieving a diagnosis, by testing the new genes, which is considered by the GLH to be significant enough to justify retesting. Relevant tests from the National Genomic Test Directory should be performed which is likely to include the genes already tested for previously.

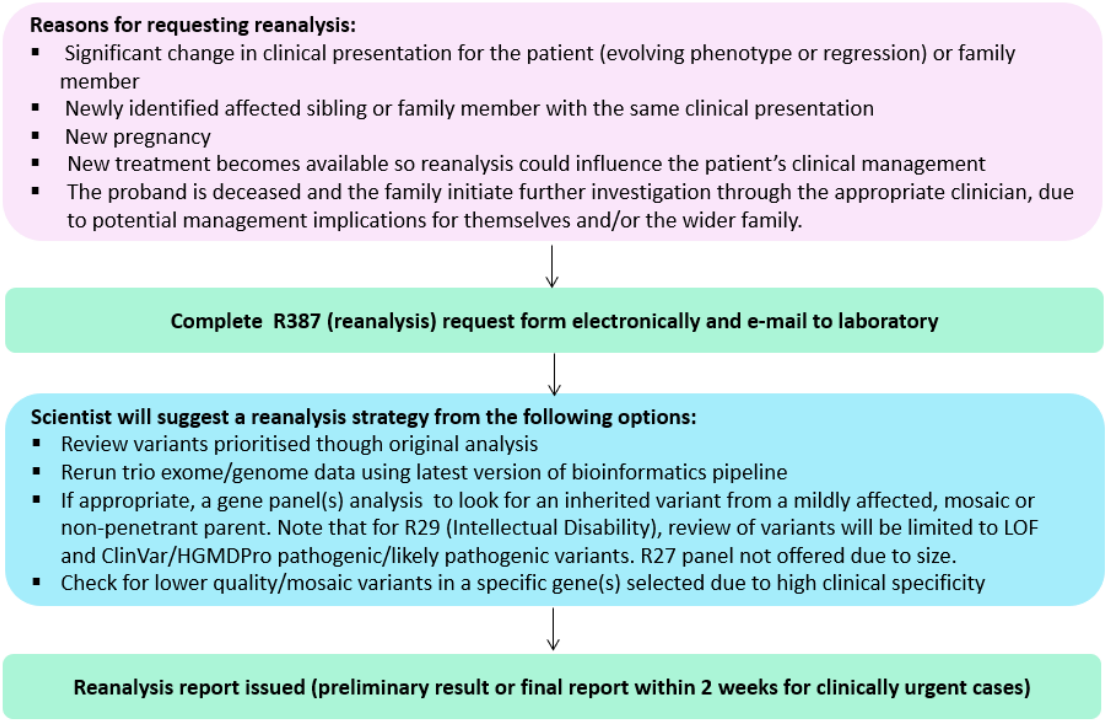
6.2 Reanalysis of R21 exome sequencing data through the R14 rapid genome sequencing service

Post-natal gene-agnostic reanalysis of R21 trio exome sequencing data (fetal anomalies with a likely genetic cause) is possible in some cases through the R14 rapid genome service for acutely unwell children at the Exeter laboratory (SW GLH). However, please note that it is not possible to undertake CNV analysis due to differences in the bioinformatics pipelines and capture technologies used therefore it may be more appropriate in eligible cases to progress as a new R14 referral. Please contact the Exeter laboratory to discuss any such requests who will in turn coordinate with the R21 sequencing laboratories.

6.3 Reanalysis of R14 exome/genome sequencing data

Reanalysis of exome or genome sequencing data for patients who have undergone R14 rapid genomic sequencing will be performed by the Exeter laboratory (SW GLH) on a case by case basis according to clinical need and the likelihood of finding a new (or additional) genetic diagnosis. Figure 4 outlines the process.

Figure 4: Pathway for reanalysis of R14 exome/genome sequencing data



6.4 Reanalysis requests where family members are considering having a child or a diagnosis could aid clinical management.

New/improved knowledge about disease-gene/variant associations and GLH bioinformatic pipeline improvements will be made over time, however systematic regular reanalysis of all historical cases is not feasible currently and diagnostic yield will be low. Where parents/family members are considering having a child or a diagnosis may aid clinical management, reanalysis requests may be considered if the initial analysis was undertaken >2 years prior to the request. **Please contact the GLH sequencing laboratory (i.e. the GLH laboratory who performed the original test) to discuss cases in advance of making a reanalysis request.**

7 Roles and Responsibilities

The overarching responsibility for approving data reanalysis lies with the GLH. Each GLH shall have a process for managing and approving such requests.

Where appropriate, reanalysis for cases within specialist areas may be outsourced to a GLH delivering relevant specialist testing.

8 Appendix

Appendix 1: Genomics England's Diagnostic Discovery analysis information and reanalysis related Q&A

Ongoing analysis of whole genome sequencing data

Diagnostic Discovery at Genomics England

Genomics England is supporting Diagnostic Discovery for participants in the 100,000 Genomes Project with ongoing analysis of the WGS data. This is in addition to research work performed using the data in the National Genomic Research Library (NGRL) by academic and commercial researchers. All putative new diagnostic findings are being returned to NHS GLHs through the Diagnostic Discovery pathway for clinical evaluation and reporting.

Genomics England are using different methods, principally based on a gene screening approach, to identify candidate variants that may represent new diagnoses. The aim is to systematically review all genes and loci included in gene panels used in the NHS GMS and additional genes added to panels as new evidence emerges. This approach efficiently identifies candidate variants in genes that have been added to applicable gene panels since the primary interpretation and mitigates other scenarios that are challenging to accommodate in automated high throughput bioinformatics pipelines (such as dual diagnoses, mosaicism, incomplete penetrance, allelic heterogeneity and technically challenging loci). Genomics England are considering all variant types but focusing initially on small variants and copy number variants. The analysis will pay particular attention to specific variants and loci that may have been impacted by limitations in the earlier pipeline, such as indels, *de novo* variants, small CNVs, non-coding regions etc.

All 100,000 Genomes Project participants with appropriate consent are included in this analysis, including participants in the Rare Disease and Cancer Programmes. This work will also be extended to include NHS GMS patients. The aim of this analysis is to identify putative diagnostic variants relevant to the primary clinical indication for the participant/patient. In accordance with the guidelines established with the Diagnostic Discovery Oversight group, additional or incidental findings will not be returned to the GLHs.

Examples of new findings that have been identified through this work to date include; numerous families with variants in established genes such as *ANKRD11*, *SHANK3* and *MECP2* that fail quality filters; >20 families with putative diagnoses in non-coding genes (*RMRP*, *RNU4ATAC*, *SNORD118, RNU7-1*); and variants in genes with new phenotypic associations, such as *H3-3A*, *MORC2* and *GEMIN5*.

Ongoing Diagnostic Discovery analysis accelerate new putative diagnostic findings for participants and patients. The data generated will also enrich the variants described in the Clinical Variant Arc (CVA), which will improve the prioritisation of variants in the GMS WGS bioinformatics pipeline through Known Pathogenic Variant Prioritisation, and help to guide future improvements for the WGS service in the GMS.

Frequently asked questions

Do I need to make a request for my patient to be included in this analysis?

No, you do not need to ask for a participant or patient to be included. All participants in the 100,000 Genomes Project are included in this ongoing analysis, provided they have not withdrawn from the project. This includes all participants from the Rare Disease programme. Analysis of pertinent germline findings will be included for participants in the Cancer programme. All patients who have whole genome sequencing through the GMS will be included.

Does a patient have to consent to their data being in the National Genomic Research Library (NGRL) to benefit from Diagnostic discovery work performed internally at Genomics England?

For participants in the 100,000 Genomes Project, all participants with valid consent for the project are included in this work. Genomics England are not able to perform this analysis and return any findings to the GLHs for participants who have withdrawn consent.

For patients who have had whole genome sequencing through the Genomic Medicine Service, all patients who have provided consent for the diagnostic test are included, regardless of whether they have given consent to share their data in the NGRL. Genomics England are able to do this as the analysis and any findings returned to the NHS GLHs for reporting are for diagnostic purposes only. However, only participants who have given consent to share their data for research purposes will benefit from other research work performed by academic researchers using the data in the NGRL.

The panels that were used in the analysis for my patient have been updated since the initial clinical analysis, are all the genes that are green on the most recent version of panels included in this work? And will new genes that are added to GMS panels over time be looked at as well?

All genes that are rated green (diagnostic-grade) in a relevant GMS panel will be analysed. This means that genes that are not included in panels used in the 100,000 Genomes Project and genes that have been added to panels since the first analysis will all be analysed. Future updates to panels will also be incorporated as new genes are added.

As knowledge grows and new pathogenic variants are detected by WGS analysis performed in the NHS, will other patients with the same variant and phenotype be identified?

Yes, the analysis approach incorporates the data stored in CVA to identify additional participants/patients with variants reported as pathogenic through NHS WGS analysis. The analysis will be repeated on an intermittent basis to incorporate newly classified variants. When putatively new diagnostic variants are identified from the ongoing analysis, other phenotypically similar patients with the same variants will be identified.

When ClinVar is updated, are new reported pathogenic variants from ClinVar looked for in all patients who have had WGS testing?

Yes, the version of ClinVar used in the Diagnostic Discovery analysis will be regularly updated and used to look for new potential diagnoses, including for participants/patients for whom the clinical analysis was performed earlier.

Does this mean that I don't need to test participants from the 100,000 Genomes Project through the GMS if they return with a question about the condition with which they joined the Project?

Yes, if the original reason for referral is unchanged, retesting through the GMS is not recommended.

If the clinical question remains unchanged, it is possible that a diagnosis could now be achieved due to updated gene panels or improved bioinformatics pipelines, and by looking at emerging candidate genes not currently included on GMS panels. However, the diagnostic yield from analysing genomes through the GMS following participation in the 100,000 Genomes Project is extremely low and the Genomics England Diagnostic Discovery analysis is likely to be a much more efficient approach to reanalysis. Reducing the number of reanalysis requests performed by NHS clinical scientists will enable more timely analysis for those patients awaiting their primary analysis.

This ongoing Diagnostic Discovery analysis is likely to result in more diagnoses overall, as a cohort-wide approach can reveal associations that a single patient analysis cannot. Diagnostic Discovery analysis also incorporates analysis for complex diagnoses, including non-coding regions and structural variants, and the analysis will be on an ongoing basis. Therefore, participants from the 100,000 Genomes Project may benefit more through Diagnostic Discovery than if they were to be reanalysed under current GMS diagnostic protocols.

Where the clinical question has changed e.g. different gene panels are required or the affected status of family members is incorrect, then it may be appropriate to request additional analysis from the GLH via a new referral to the GMS WGS pathway. However, in some of these scenarios, it is still possible that a diagnosis may be identified through Diagnostic Discovery, for example by accounting for incomplete penetrance or if new clinical data are provided to Genomics England by the referring clinician.

However, it should be noted that Genomics England's Diagnostic Discovery analysis is NOT intended as a replacement for clinical reanalysis/additional analysis and the NHS Reanalysis Guidance should be followed, especially in any clinically urgent scenarios.

Does this mean that I don't need to request reanalysis by the GLH for my patient?

If the clinical question remains unchanged and reanalysis is being requested as sufficient time has passed since the primary analysis for there to have been significant changes to knowledge, gene panel content and bioinformatic pipelines, then it is not necessary to request reanalysis from the GLH. Diagnostic Discovery analysis aims to look for pathogenic variants that are revealed as a result of this improved knowledge over time and provides an efficient approach to identifying new findings.

However, where additional analysis is required due to a requirement for different analysis parameters (such as more gene panels, using incomplete penetrance settings and/or a change to affected status), then this can be requested via the GLH for patients who have undergone primary WGS analysis in the GMS. See the NHS Reanalysis Guidance for further details.

How do I update the clinical information for my patient recruited to the 100,000 Genomes Project?

Updating the clinical data for participants recruited to the 100,000 Genomes Project will be beneficial to the Genomics England Diagnostic Discovery analysis. This could be a more recent or detailed phenotypic description, or a change to the affected status of family members. Updating the clinical data for a family can help Diagnostic Discovery support scenarios such as analysis with additional gene panels or assuming incomplete penetrance. If you wish to update the clinical data for a patient recruited to the 100,000 Genomes Project, please submit a Service Desk ticket to Genomics England clearly describing the changes. Please specify the additional Human Phenotype Ontology (HPO) terms to be added or describe other changes. Please see https://hpo.jax.org/app/ for more information on the Human Phenotype Ontology.

The updated data will be made available to the Genomics England Diagnostic Discovery analysis work and in the NGRL. The clinical data used during the primary clinical interpretation of the 100,000 Genomes Project data will remain unchanged.

9 References

Ellard. (2020). *ACGS Best Practice Guidelines for Variant Classification*. Retrieved from https://www.acgs.uk.com/quality/best-practice-guidelines/

Ellingford, J. M. (2018). *Whole genome sequencing enables definitive diagnosis of Cystic Fibrosis and Primary Ciliary Dyskinesia*. Retrieved from https://www.biorxiv.org/content/10.1101/438838v1

Evans, D. (2018). A Dominantly Inherited 5′ UTR Variant Causing Methylation-Associated Silencing of BRCA1 as a Cause of Breast and Ovarian Cancer. *Am. J. Hum. Genet, 103*, 213-220. Retrieved from https://www.sciencedirect.com/science/article/pii/S0002929718302295

Flanagan, S. (2013). Next-Generation Sequencing Reveals Deep Intronic Cryptic ABCC8 and HADH Splicing Founder Mutations Causing Hyerinsulinism by Pseudoexon Activation - ScienceDirect. *Am. J. Hum Genet., 92*(1), 131-136. Retrieved from https://www.sciencedirect.com/science/article/pii/S0002929712006301

Guidelines for Rare Disease Whole Genome Sequencing & Next Generation Sequencing Panel Interpretation & Reporting. (n.d.). Retrieved from https://future.nhs.uk/NHSgenomics/viewdocument?docid=154355013&done=DOCCreated1&fid=12981968

Kvikstad, E. (2018). A high throughput screen for active human transposable elements. *BMC Genomics, 19:*(115). Retrieved from https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4485-4

NHS Diagnostic Discovery Guidance. (n.d.). Retrieved from NHS Futures : https://future.nhs.uk/NHSgenomics/view?objectId=180872901

R387 Reanalysis request form. (n.d.). Retrieved from NHS Futures : https://future.nhs.uk/NHSgenomics/view?objectId=156197061

Richards. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine, 17*, 406-423. Retrieved from https://www.nature.com/articles/gim201530?ux=07df2189-4e01-4c08-8ef3-5619cff0ca61&ux2=3739b439-66b5-4bf5-921e-0916eef236a7&ux3=&uxconf=Y

Weedon, M. (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nature Genetics, 46*, 61-64. Retrieved from https://www.nature.com/articles/ng.2826