Supplemental Online Content

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods. Supplemental Methods and Translated Copies of Informed Consent, Clinical Utility, Outcome, and Requisition Forms

eForm 1:

<u>**Information and Consent form for Next Generation Sequencing Based Genetic Testing in a neonate suspected of genetic disease**</u>

What is Next Generation Sequencing?

These tests use technology that allows simultaneous testing of many genes associated with function and disease. These techniques have become an important tool in the diagnosis of diseases and genetic syndromes.
Whole genome sequencing (WGS) A test of the entire sequence of the human genome (about three billion letters), which includes about 20,000 genes including protein coding and non-coding intervals within and between them. Variants in the DNA sequence that encode for protein may cause genetic disease and are called: "disease causing genetic variant" or a "pathogenic variant". Identifying such variants is important for the treatment and allows genetic counseling for family members.

Explanation of the test technique:
To perform the test, a DNA samples from the subject and his parents (a trio model) are sent to an authorized laboratory that performs the next generation sequencing test. The resulting genetic sequence is decoded by a team of expert bioinformaticians and is compares to the normal genetic sequence. The decoding is based on the accumulated updated information from the medical literature and databases (including the database of Israeli patients), detailed report of the patient's medical condition, his family members and an accurate information of biological family relations.
The test results detail the significant variants detected which is deciphered by geneticists for clinical significance. Time to result is about 14 working days.
It is sometimes necessary to validate variants by another test method such as Sanger sequencing (which is a separate test and is not included in the next generation sequencing).
Sometimes further investigation and DNA testing of family members is needed to determine if the variants are indeed significant to the reported medical condition (a separate test that is not included in the next generation sequencing). The duration of this test is several weeks.
Test results:
There are several possible test results:
1. The found genetic variant is the cause of the disease in the subject.
2. No genetic variant was found that explains the disease in the subject.
3. A genetic variant has been identified that leads to a disease different from that reported in the subject.
4. A genetic variant was found with unclear clinical significance.

<u>**In general, we report variants that are classified as significant for genetic diseases related to the clinical condition for which the test was ordered.**</u>
Variants related to other diseases in the patient and family members will not be analysed and if will be discovered incidentally, will not be reported except at the discretion of the referring geneticist and subject to the consent of the subject (see below).
A summary report of the test will be sent to the referring physician who will deliver the results to the patient in a formal genetic counseling session.
The raw digital data file containing the DNA sequences (FASTQ files) will be saved at the sequencing laboratory (TASMC) for two years and will be transferred to the referring genetic institute and later on to the Ministry of Health genetic database.

Disadvantages and limitations of the test:
1. For some of the variants that are identified there is no conclusive information regading their clinical significance. In some cases, parental DNA testing will be required to understand the meaning of such variants (these are included in the trio test) or additional family members. In some cases the classification of the variant is not feasible due to limited information regarding the variant in the medical literature at the time of the analysis.
2. There may be differences in bioinformatic analysis between different laboratories and the bioinformatics tools, and information in the literature change and evolve from time to time which may affect interpretation and outcomes. Thus there may be differences in the decoding and interpretation of the same raw data in different laboratories and at different times. It is possible to request (at an additional cost) a reanalysis of the raw data (FASTQ files) in the future. The request is the responsibility of the patient and his/her family members.
3. As a rule, the Genetic Institute does not conduct a re-examination (revision) of the data from time to time on its own initiative. This may be performed at the request of the subject and family members, after repeat genetic counseling and at an additional cost.

<u>Informed Consent for whole Genome Sequencing (proband)</u>

Full name: _____

ID No: _____

Address: _____

Cell phone no.: _____

Purpose of the test:
Identification of genetic variants in the DNA sequence, the causes of the disease / genetic syndrome / clinical signs in the main subject. Testing of other family members is only intended to assist in the analysis of the results in the main subject.

Receiving the results: the test result will be sent to the referring geneticist and will be delivered to the patient in a genetic counseling session.

St**orage of the test results: According to the instruction from the Ministry of Health,** the performing genetic laboratory will store the raw data (FASTQ files) for two years. The raw data is used continuously as a database for bioinformatic analysis.

- I confirm that I have read the informed consent form for next generation testing, I have received appropriate counseling, I have been given the opportunity to ask questions, and I request to perform the test. I am aware that the test results will be sent to the referring geneticist and that I will receive the test results and their consequences in a formal genetic counseling.
- I am aware that there is a possibility that the test will not yield a genetic explanation for the disease or the condition for which I was referred. Even if no genetic explanation is found, there may still be a genetic / hereditary basis for the disease or the condition for which my child was referred, which may have an effect on me or my family members, including increased risks of recurrence in the offspring.
- **Secondary Findings:** The American College of Medical Genetics and Genomics (ACMGG) has established a list of genes associated with high-risk morbidity for which early detection can significantly reduce morbidity and mortality by established interventions (ACMG SF V3.1<u>). As a rule, such variants will be reported in the subject, unless the subject has explicitly requested that they should not be reported. In addition, minors will not receive heterozygote variants in genes associated with an increased risk of adult onset cancer</u> (such as *BRCA1, BRCA2, MSH2, MSH6, MLH1, PMS2,* etc.).

☐ I request **not** to receive a report of **secondary variants** that are not related to the condition for which the test is being conducted.
Signature (if requesting not to report): _____

- **Incidental findings:** Occasionally during the analysis, incidentally and unintentionally, variants of medical significance are identified in the main subject or other family members, that are unrelated to the disease/condition for which the next generation testing was ordered. Such variants may include predisposition for adult onset diseases, various types of cancer, and other hereditary diseases. The test is not intended to detect such variants and they may be discovered only at random. <u>Generally, these variants will be reported to the subject at the discretion of the clinical geneticist, unless the subject has specifically requested that they should not be reported.</u> The information may be delivered to the referring geneticist and upon their discretion the subject and/or his family members/parents/guardian will be invited for genetic counseling to conveyed the results.

☐ I request not to be notified of **incidental findings** that are not related to the condition for which the test is being conducted.
Signature (if requesting not to report): _____

- **Carriers for recessive diseases:** the test does not serve as an expanded genetic carrier testing for identification of carriers for recessive diseases. Therefore, as a rule, variants relating to carriers of recessive diseases are not reported.
- **Genetic variants of unknown significance**: Further investigations of such findings will be pursed at the discretion of the treating geneticist. If relevant, the doctor / genetic counselor will list the testing options upon delivery of test results.
- The test results will be delivered to the family members within a genetic counseling session and only after their explicit consent to receive the results.
- As a rule, the Genetic Institute does not conduct a re-examination (revision) of the data from time to time on its own initiative, but at the request and initiative of the examinee as part of genetic counseling and at an additional fee. It is advisable to be updated on the need for revision from time to time.
- I approve the storage of the data files in the database of the Genetic Institute performing the test, the referring genetic institute and anonymously in the database of the Ministry of Health (Department of Genetics).

Signature of patient/parent/guardian _____ date: _____

When relevant:
Name and ID of Legal Guardian/Parent: _____
(please enclose guardianship certificate)

Signature:_____Date: _____

Name and signature of Medical geneticist/Genetic counselor_____

_____

<u>Informed Consent for whole Genome Sequencing (mother of the proband)</u>

Full name: _____

ID No: _____

Address: _____

Cell phone no.: _____

Type of test:
☐ **Whole genome sequencing**

Purpose of the test:
Identification of genetic variants in the DNA sequence, the causes of the disease / genetic syndrome / clinical signs in the main subject. Testing of other family members is only intended to assist in the analysis of the results in the main subject.

Receiving the results: the test result will be sent to the referring geneticist and will be delivered to the patient in a genetic counseling session.

Storage of the test results: According to the instruction from the Ministry of Health, the performing genetic laboratory will store the raw data (FASTQ files) for two years. The raw data is used continuously as a database for bioinformatic analysis.

- I confirm that I have read the informed consent form for next generation testing, I have received appropriate counseling, I have been given the opportunity to ask questions, and I request to perform the test. I am aware that the test results will be sent to the referring geneticist and that I will receive the test results and their consequences in a formal genetic counseling.
- I am aware that there is a possibility that the test will not yield a genetic explanation for the disease or the condition for which I was referred. Even if no genetic explanation is found, there may still be a genetic / hereditary basis for the disease or the condition for which my child was referred, which may have an effect on me or my family members, including increased risks of recurrence in the offspring.
- **Secondary Findings:** The American College of Medical Genetics and Genomics (ACMGG) has established a list of genes associated with high-risk morbidity for which early detection can significantly reduce morbidity and mortality by established interventions (ACMG SF V3.1<u>). As a rule, such variants will be reported in the subject, unless the subject has explicitly requested that they should not be reported. In addition, minors will not receive heterozygote variants in genes associated with an increased risk of adult onset cancer</u> (such as *BRCA1, BRCA2, MSH2, MSH6, MLH1, PMS2*, etc.).

☐ I request **<u>not</u>** to receive a report of **secondary variants** that are not related to the condition for which the test is being conducted.
Signature (if requesting not to report): _____

- **Incidental findings:** Occasionally during the analysis, incidentally and unintentionally, variants of medical significance are identified in the main subject or other family members, that are unrelated to the disease/condition for which the next generation testing was ordered. Such variants may include predisposition for adult onset diseases, various types of cancer, and other hereditary diseases. The test is not intended to detect such variants and they may be discovered only at random. <u>Generally, these variants will not be reported in the parents of the main subject.</u>

- **Carriers for recessive diseases:** the test does not serve as an expanded genetic carrier testing for identification of carriers for recessive diseases. Therefore, <u>as a rule, variants relating to carriers of recessive diseases are not reported.</u>
- **Genetic variants of unknown significance**: Further investigations of such findings will be pursed at the discretion of the treating geneticist. If relevant, the doctor / genetic counselor will list the testing options upon delivery of test results.
- The test results will be delivered to the family members within a genetic counseling session and only after their explicit consent to receive the results.
- As a rule, the Genetic Institute does not conduct a re-examination (revision) of the data from time to time on its own initiative, but at the request and initiative of the examinee as part of genetic counseling and at an additional fee. It is advisable to be updated on the need for revision from time to time.
- I approve the storage of the data files in the database of the Genetic Institute performing the test, the referring genetic institute and anonymously in the database of the Ministry of Health (Department of Genetics).

Signature of patient/parent/guardian _____**date:** _____

When relevant:
Name and ID of Legal Guardian/Parent: _____
(please enclose guardianship certificate)

Signature:_____Date: _____

Name and signature of Medical geneticist/Genetic counselor_____
_____

<u>Informed Consent for whole Genome Sequencing (father of the proband)</u>

Full name: _____

ID No: _____

Address: _____

Cell phone no.: _____

Type of test:
☐ **Whole genome sequencing**

Purpose of the test:
Identification of genetic variants in the DNA sequence, the causes of the disease / genetic syndrome / clinical signs in the main subject. Testing of other family members is only intended to assist in the analysis of the results in the main subject.

Receiving the results: the test result will be sent to the referring geneticist and will be delivered to the patient in a genetic counseling session.

Storage of the test results: According to the instruction from the Ministry of Health, the performing genetic laboratory will store the raw data (FASTQ files) for two years. The raw data is used continuously as a database for bioinformatic analysis.

- I confirm that I have read the informed consent form for next generation testing, I have received appropriate counseling, I have been given the opportunity to ask questions, and I request to perform the test. I am aware that the test results will be sent to the referring geneticist and that I will receive the test results and their consequences in a formal genetic counseling.
- I am aware that there is a possibility that the test will not yield a genetic explanation for the disease or the condition for which I was referred. Even if no genetic explanation is found, there may still be a genetic / hereditary basis for the disease or the condition for which my child was referred, which may have an effect on me or my family members, including increased risks of recurrence in the offspring.
- **Secondary Findings:** The American College of Medical Genetics and Genomics (ACMGG) has established a list of genes associated with high-risk morbidity for which early detection can significantly reduce morbidity and mortality by established interventions (ACMG SF V3.1<u>). As a rule, such variants will be reported in the subject, unless the subject has explicitly requested that they should not be reported. In addition, minors will not receive heterozygote variants in genes associated with an increased risk of adult onset cancer</u> (such as *BRCA1, BRCA2, MSH2, MSH6, MLH1, PMS2*, etc.).

☐ I request **<u>not</u>** to receive a report of **secondary variants** that are not related to the condition for which the test is being conducted.
Signature (if requesting not to report): _____

- **Incidental findings:** Occasionally during the analysis, incidentally and unintentionally, variants of medical significance are identified in the main subject or other family members, that are unrelated to the disease/condition for which the next generation testing was ordered. Such variants may include predisposition for adult onset diseases, various types of cancer, and other hereditary diseases. The test is not intended to detect such variants and they may be discovered only at random. <u>Generally, these variants will not be reported in the parents of the main subject.</u>

- **Carriers for recessive diseases:** the test does not serve as an expanded genetic carrier testing for identification of carriers for recessive diseases. Therefore, <u>as a rule, variants relating to carriers of recessive diseases are not reported.</u>
- **Genetic variants of unknown significance**: Further investigations of such findings will be pursed at the discretion of the treating geneticist. If relevant, the doctor / genetic counselor will list the testing options upon delivery of test results.
- The test results will be delivered to the family members within a genetic counseling session and only after their explicit consent to receive the results.
- As a rule, the Genetic Institute does not conduct a re-examination (revision) of the data from time to time on its own initiative, but at the request and initiative of the examinee as part of genetic counseling and at an additional fee. It is advisable to be updated on the need for revision from time to time.
- I approve the storage of the data files in the database of the Genetic Institute performing the test, the referring genetic institute and anonymously in the database of the Ministry of Health (Department of Genetics).

Signature of patient/parent/guardian _____**date:** _____

When relevant:
Name and ID of Legal Guardian/Parent: _____
(please enclose guardianship certificate)

Signature:_____Date: _____

Name and signature of Medical geneticist/Genetic counselor_____

eForm 2: Clinical utility questionnaire

Study ID: _____

The Clinician-reported Genetic testing Utility InDEx (C-GUIDE)™

The Clinician-reported Genetic testing Utility InDEx (C-GUIDE)™ aims to capture the clinical utility of genetic testing once results are disclosed, from the perspective of the ordering clinician.
Thinking about the result you just disclosed, please complete the following:

N.B. If you disclosed multiple results, please complete the C-GUIDE once for each result disclosed.

Item	Response Options
The genetic testing that my patient had…	
1. Provided a genetic explanation for my patient's health condition	☐ Provided a COMPLETE genetic explanation ☐ Provided a PARTIAL genetic explanation ☐ Provided a POSSIBLE genetic explanation ☐ Provided NO genetic explanation
2. Reduced the likelihood of other potential diagnoses in my differential	☐ COMPLETELY REDUCED the likelihood of other potential diagnoses in my differential ☐ PARTIALLY REDUCED the likelihood of other potential diagnoses in my differential ☐ DID NOT REDUCE the likelihood of other potential diagnoses in my differential ☐ Not applicable
3. Provided information about the natural history of or medical issues associated with my patient's condition	☐ Provided SIGNIFICANT information about the natural history of or medical issues associated with my patient's condition ☐ Provided SOME information about the natural history of or medical issues associated with my patient's condition ☐ Provided NO information about the natural history of or medical issues associated with my patient's condition
4. Indicated that further testing to identify a genetic diagnosis can be avoided	☐ Indicated that further testing to identify a genetic diagnosis CAN BE AVOIDED ☐ Indicated that further testing to identify a genetic diagnosis MAY STILL BE REQUIRED, now or in the future
5. Prompted a referral or investigation for the purpose of surveillance or monitoring that would not have been prompted on clinical grounds*	☐ PROMPTED a referral or investigation for surveillance/monitoring ☐ PROMPTED a referral or investigation for surveillance/monitoring that may not be necessary (e.g. variant of uncertain significance) ☐ DID NOT PROMPT a referral/investigation for surveillance/monitoring
6. Provided information to guide medication management*	☐ GUIDED current medication management ☐ MAY GUIDE medication management in the future ☐ DID NOT PROVIDE information that would guide medication management, now or in the future
7. Provided information about surgical management*	☐ Enabled me to DISCUSS or OFFER a surgical option ☐ Enabled me to AVOID a surgical option ☐ A surgical option is NOT RELEVANT at this time or NOT RELATED to the genetic test results
8. Provided information about a contraindicated behaviour (e.g. competitive sports)*	☐ ENABLED me to provide information about a contraindicated behaviour ☐ Information about a contraindicated behaviour is NOT RELVANT at this time
9. Clarified potential health risks for my <u>patient's family</u>*	☐ YES – PROMPTED genetic testing and downstream clinical investigations for my patient's family ☐ YES – PROMPTED genetic testing for my patient's family ☐ NO - DID NOT PROMPT testing or downstream clinical investigations for my patient's family ☐ Cannot be determined (e.g. variant of uncertain significance, unknown whether family member(s) pursued testing)
10. Generated psychosocial effect for my patient <u>or</u> his/her family*	☐ SIGNIFICANT psychosocial effect was experienced ☐ MODERATE psychosocial effect was experienced ☐ NO psychosocial effect was experienced ☐ Cannot be determined

11. Prompted better care for my patient <u>or</u> his/her family overall	☐ PROMPTED BETTER care ☐ PROMPTED SOMEWHAT BETTER care ☐ Unsure if prompted better care

eForm 3: Outcome questionnaire

Case no. _____

Date _____

<div align="center">

Clinical Report Patient Outcomes

</div>

(To be filled out by the referring neonatologist)

Effects on hospitalization course	
Did the patient got discharged from the hospital since he undergo the genetic testing (WGS)?	**Yes** ☐ **No** ☐
Was the patient discharged before receiving the results of the genetic test?	**Yes** ☐ **No** ☐
Demise	
Is the patient deceased?	**Yes** ☐ **No** ☐
Date of demise	_____
Cause for demise	_____
Was the patient diagnosed with a genetic disease around the time of his demise?	**Yes** ☐ **No** ☐
Was the genetic diagnosis done using genome sequencing (WGS)?	**Yes** ☐ **No** ☐
Did the patient die before receiving the results of the WGS testing?	**Yes** ☐ **No** ☐

Secondary Variants

Was/were secondary variant(s) also disclosed? If yes, please complete the following once for all secondary results disclosed:

N.B. For the purpose of this index, secondary variants include <u>medically actionable variants unrelated to the indication for testing</u>.

Item	Response options
The genetic testing that my patient ha…	
1. Prompted a referral or investigation for the purpose of surveillance or monitoring that would not have been prompted on clinical grounds	☐ PROMPTED a referral or investigation for surveillance/monitoring ☐ PROMPTED a referral or investigation for surveillance/monitoring that may not be necessary (e.g. variant of uncertain significance) ☐ DID NOT PROMPT a referral/investigation for surveillance/monitoring
2. Provided information to guide management	☐ GUIDED current management ☐ MAY GUIDE management in the future ☐ DID NOT PROVIDE information that would guide management, now or in the future
3. Provided information about the presence or absence of reproductive risk for my <u>patient's family</u>	☐ Provided information about the presence or absence of reproductive risk that is RELEVANT to my patient's family at this time ☐ Provided information about the presence or absence of reproductive risk that MAY BE RELEVANT to my patient's family in the future ☐ Cannot be determined (e.g. variant of uncertain significance, did not provide information)
4. Generated psychosocial effect for my patient <u>or</u> his/her family*	☐ SIGNIFICANT psychosocial effect was experienced ☐ MODERATE psychosocial effect was experienced ☐ NO psychosocial effect was experienced ☐ Cannot be determined

eFORM 4: Candidate requisition form (RF).

The RF was completed for each candidate and emailed to the committee in the MOH for review.

Candidate requisition form (HPO terms)

Date at referral (dd/mm/yyyy) _____

Age at symptom onset _____days/ weeks

Suggested differential diagnosis _____

Please mark one criterion:

1. Severe resistant convulsive disease

2. Encephalopathy

3. Congenital severe persistent abnormality of muscle tone

4. Two separate major anomalies

5. Single otherwise rare abnormality[1]

6. Structural brain malformation

7. Unstable metabolic or endocrine abnormality

8. Congenital heart failure or cardiomyopathy

Please circle appropriate terms (one or more):

1. Abnormality of prenatal development or birth (HP:0001197)
2. Abnormal heart morphology (HP:0001627)
3. Cardiac conduction abnormality (HP:0031546)
4. Abnormality of the skeletal system (includes arthrogryposis) (HP:0000924)
5. Abnormality of the skin (HP:0000951)
6. Abnormality of metabolism/homeostasis (HP:0001939)
7. Abnormality of the endocrine system (includes ambiguous genitalia) (HP:0000818)
8. Abnormality of the gastrointestinal tract (HP:0011024)

9. Abnormality of blood and blood-forming tissues (HP:0001871)
10. Hepatic failure (HP:0001399)
11. Brain imaging abnormality (HP:0410263)
12. Seizure (HP:0001250)
13. Generalized hypotonia (HP:0001290)
14. Encephalopathy (HP: 0001298)
15. Abnormal cerebral white matter morphology (HP: HP:0002500)
16. Abnormal renal morphology (HP:0012210)
17. Renal insufficiency (HP:0000083)
18. Abnormality of the respiratory system (HP:0002086)

HPO		HPO	

Are parents blood related	No	YES by decent
Ethnic background	Mother:	Father:
Family history of same phenotype	No	Yes, Specify
Family history of genetic disease	No	Yes, Specify
Did the parents pursue genetic testing during pregnancy (NIPT, karyotype, CMA, exome sequencing, carrier screening, etc.)	No	Yes, Specify test/s and results_____

eTable 1. Study Inclusion and Exclusion Criteria

One Primary inclusion criteria was selected for each candidate and listed on an excel sheet emailed with the RF and medical summary to the MOH.

	Prerequisite criteria for eligibility for rtGS	
	Age ≤90 days	
	Both parents available for genome sequencing	
	Written informed consent by both parents	
	Primary category inclusion criteria	
1	Primary neurologic phenotype:	
	Severe resistant convulsive disease	
	Encephalopathy	
	Congenital severe persistent abnormality of muscle tone	
2	Two separate major anomalies	
3	Single otherwise rare abnormality[1]	
4	Structural brain malformation	
5	Unstable metabolic or endocrine abnormality	
6	Congenital heart failure or cardiomyopathy	
	Secondary category inclusion criteria (HPO terms)	
1	Abnormality of prenatal development or birth	HP:0001197
2	Abnormal heart morphology	HP:0001627
3	Cardiac conduction abnormality	HP:0031546
4	Abnormality of the skeletal system (includes arthrogryposis)	HP:0000924
5	Abnormality of the skin	HP:0000951
6	Abnormality of metabolism/homeostasis	HP:0001939
7	Abnormality of the endocrine system (includes ambiguous genitalia)	HP:0000818
8	Abnormality of the gastrointestinal tract	HP:0011024
9	Abnormality of blood and blood-forming tissues	HP:0001871
10	Hepatic failure	HP:0001399
11	Brain imaging abnormality	HP:0410263
12	Seizure	HP:0001250
13	Generalized hypotonia	HP:0001290
14	Encephalopathy	HP:0001298
15	Abnormal cerebral white matter morphology	HP:0002500
16	Abnormal renal morphology	HP:0012210
17	Renal insufficiency	HP:0000083
18	Abnormality of the respiratory system	HP:0002086
	Exclusion criteria	
1	Acquired causes as sole explanation of phenotype	
2	Toxicity	
3	Infection/sepsis	
4	Prematurity	
5	Hypoxic ischemic encephalopathy	

[1]e.g., skeletal dysplasia, cardiac arrhythmia, liver failure, etc. Abbreviations: rtGS, rapid trio genome sequencing; HPO, Human Phenotype Ontology.

eMethods 2. Statistical Analyses

All statistical analyses for this study were conducted using R.[1]

Fisher's exact test[2] was selected for categorical comparisons in the analysis due to small sample sizes in specific categories, which could potentially violate the assumptions of the Chi-square test, which requires all expected values to be greater than 1 and at least 20% of the expected values to be greater than 5[3], a condition that was not met in some categories. Fisher's exact tests were used for all categorical comparisons, regardless of sample size, to maintain consistency in the statistical approach.

For continuous variables, the Kruskal-Wallis[4] test was employed. This non-parametric method was preferred over ANOVA as it does not assume a normal data distribution, making it a more robust choice for the dataset. The Kruskal-Wallis test is beneficial when comparing more than two groups, as in this study, with three diagnosis groups: diagnosed, possible diagnosis, and undiagnosed.

To delve deeper into the relationships between variables and the diagnosis status, a logistic regression (LR) model was constructed. The LR model was chosen for its ability to handle categorical outcome variables and its robustness in analysing relationships between variables. Only variables with complete data for all 130 probands were included in the LR model to ensure data integrity and completeness. The Ministry of Health terms were excluded from the model due to their lack of significance in the Fisher's exact tests and their overlap with the Human Phenotype Ontology (HPO) terms. This approach aimed to create a more parsimonious model by reducing redundancy and focusing on the most influential variables.

In the LR analysis, several models were considered, including stepwise models[5] based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC),[6] a model using backward elimination[5], and models using Recursive Feature Elimination (RFE) with Random Forest.[7] Each of these models has its strengths and weaknesses, and they can provide different insights depending on the specific characteristics of the data.

The stepwise models (AIC and BIC) are automated methods that add or remove predictors based on their statistical significance, to optimize the model's fit to the data. The backward elimination model starts with all predictors and removes the least significant ones one by one until all remaining predictors are statistically significant. The RFE models use a machine learning algorithm (Random Forest) to rank the predictors based on their importance and eliminate the least important predictors.

However, despite the sophistication of these methods, the original model, which included all predictors, was found to be the best based on several performance metrics, including area under the curve (AUC), precision, recall, and F1 score. One plausible reason for this is the small sample size of our study, which can make the more complex models prone to overfitting. It is important to note that the 15 cases with variants of uncertain significance (VUS) were not included in the LR models. This decision was made because the diagnostic status of these cases is still uncertain and including them in the models could potentially introduce noise and bias. However, we did use the model built on the other cases to predict the diagnostic status of the 15 VUS cases. This approach allows us to use all available data while still maintaining the integrity of the models.

Logistic Regression (LR) Model

An LR model was built to investigate further the relationships between the variables and the diagnosis status. Several models were considered, and the original model (Model A) was the best based on the AUC, precision, recall, and F1 score metrics. The performance of the different models is summarized in the table below.

Model	AUC	Precision	Recall	F1 score
A: Original	0.856	0.848	0.812	0.830
B: Stepwise (AIC)	0.832	0.848	0.767	0.806
C: Stepwise (BIC)	0.644	0.875	0.140	0.241
D: Backward elimination	0.832	0.767	0.660	0.710
E: RFE_Random_Forset_1	0.759	0.783	0.818	0.800
F: Recursive Feature Elimination_2	0.731	0.750	0.300	0.429

The original model (Model A) was the best based on the AUC, precision, recall, and F1 score metrics.

The receiver operating characteristic (ROC) curve (Supplementary Figure S2) and confusion matrix (Supplementary Figure S3) portray the model's performance. The ROC curve illustrates the trade-off between correctly identifying positive cases and mistakenly identifying negative cases as positive. The AUC provides a single metric for model quality—the closer it is to 1, the better the model. The confusion matrix visualizes the accuracy of our model by showing correct and incorrect predictions.
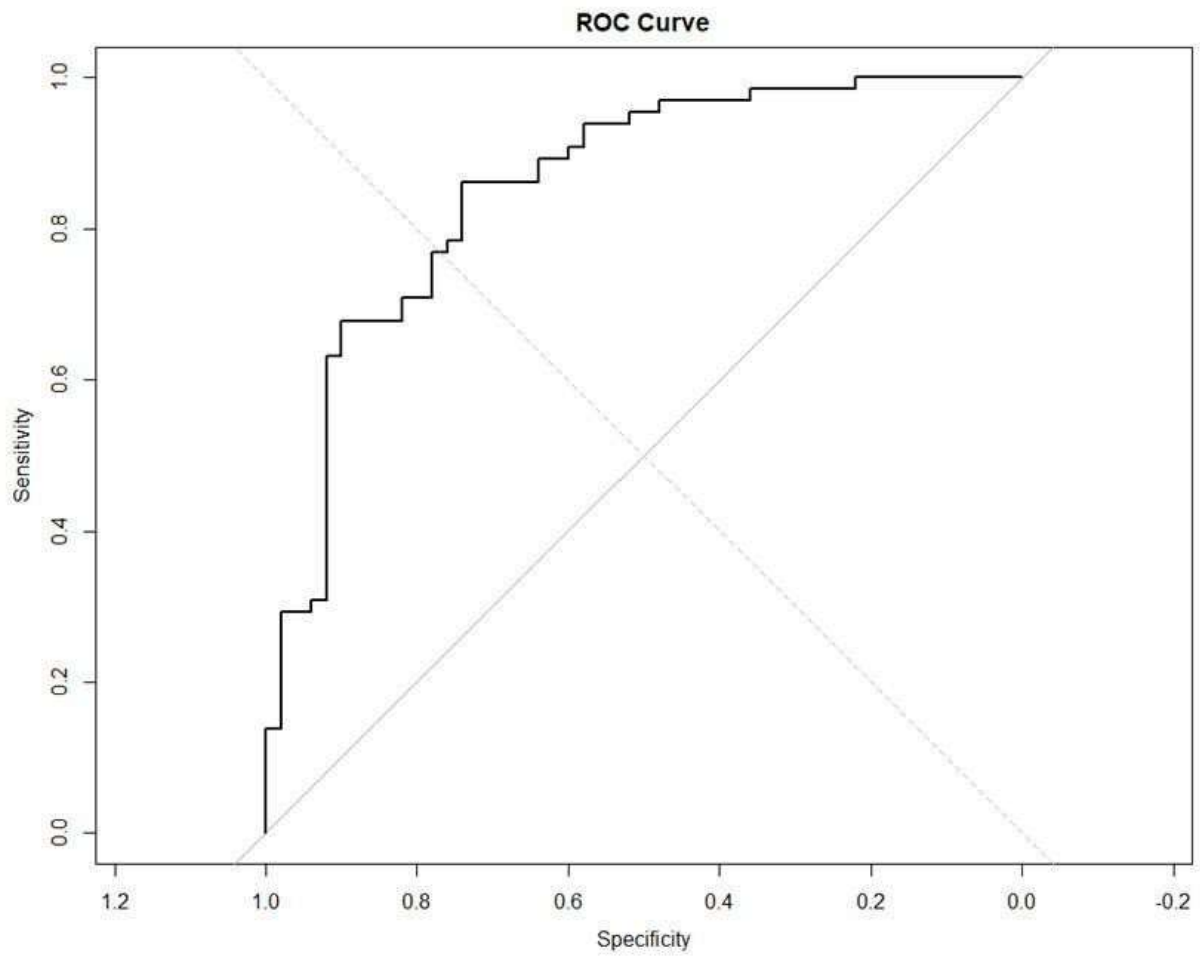
The predictive model's receiver operating characteristic (ROC) curve. Area under the curve (AUC) indicates model performance (0.856).
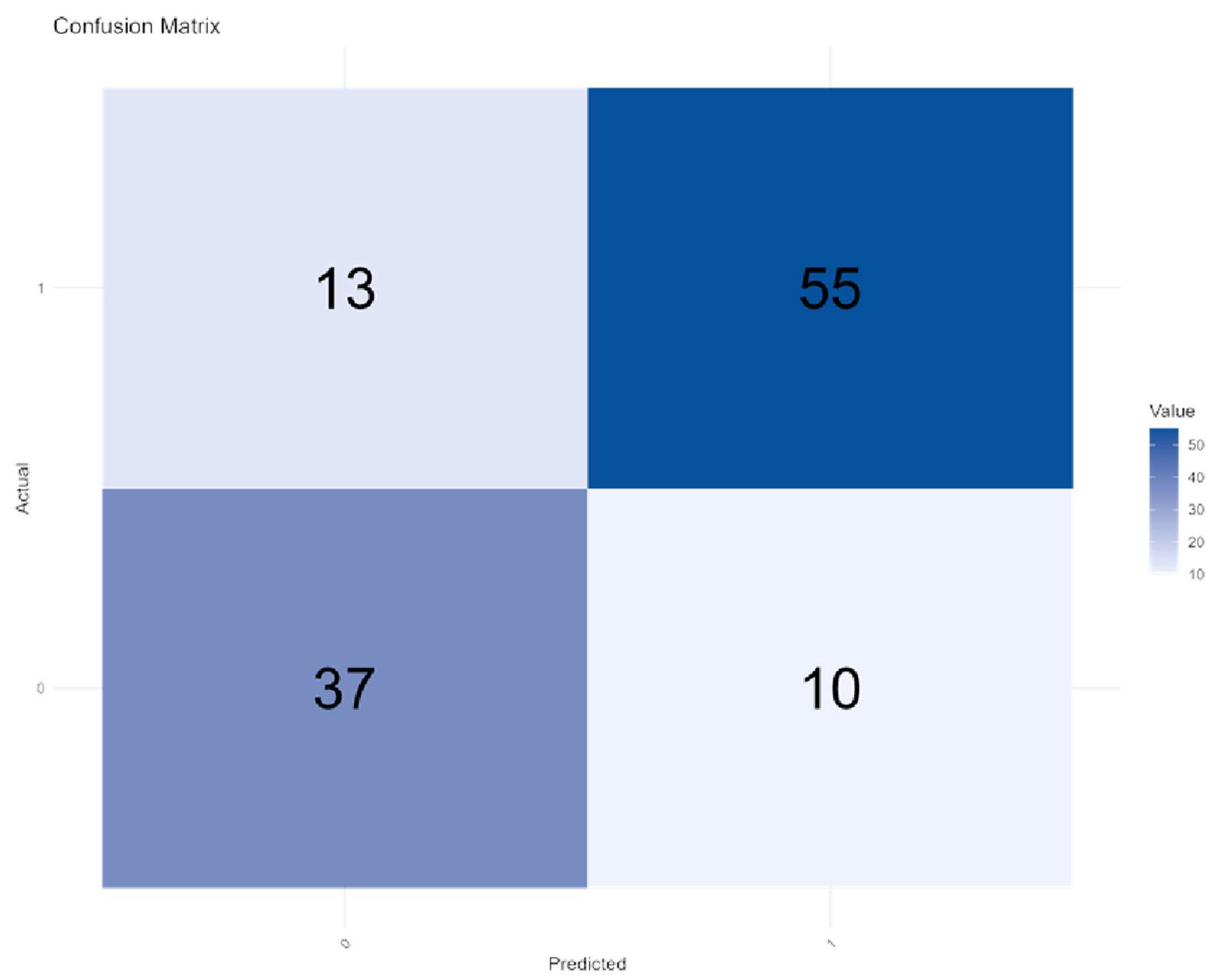
A confusion matrix representation of our model's predictive output. The matrix maps out true positives/negatives (correct predictions for diagnosed/undiagnosed cases) and false positives/negatives (incorrect predictions), clearly showing correct and incorrect classifications across all instances.

It is important to note that the 15 cases with variants of uncertain significance (VUS) were not included in the LR models. This decision was made because the diagnostic status of these cases is still uncertain and including them in the models could potentially introduce noise and bias. However, we did use the model built on the other cases to predict the diagnostic status of the 15 VUS cases. This approach allows us to use all available data while still maintaining the integrity of the models.

eReferences.

1. R Core Team. R: A language and environment for statistical computing. 2018. https://www.R-project.org/ (accessed June 22, 2023).

2. Fisher RA. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* 1922; **85**(1): 87-94.

3. Yates D, Moore DS, McCabe GP. Practice of Statistics: Advanced Placement Edition: 1st (first) Edition: W H Freeman & Co.; 1999.

4. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* 1952; **47**(260): 583-621.

5. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine* 2008; **3**(1): 17.

6. Stoica P, Selen Y. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine* 2004; **21**(4): 36-47.

7. Misra P, Yadav AS. Improving the classification accuracy using recursive feature elimination with cross-validation. *Int J Emerg Technol* 2020; **11**(3): 659-65.

eAppendix. Supplemental Results

Genome Sequencing Performance

Illumina DNA PCR-Free Prep kit was used. Sequencing was done on Novaseq 6000 150 bp-paired-end reads running kits, aiming for average coverage of 30X per sample. FASTQ files were analyzed first using TruSight Software Suite (Illumina), and second analysis was done using Franklin genetic analysis platform (Genoox, Tel Aviv, Israel). FASTQ files were aligned to hg19/GRCh37. Analysis included single-nucleotide polymorphisms (SNPs), small indels, copy-number variation and repeat expansion analysis for specific genes. Number of reads per sample >200,000,000. Average coverage depth was >30x (ranging X20-x80), number of SNPs (excluding low quality) per sample: ~5,500,000.

rtGS result predictors variables

In the LR analysis, some variables that were not significant in Fisher's exact tests emerged as important features. For instance, hepatic failure, seizure, generalized hypotonia, abnormality of the endocrine system, and abnormal renal morphology all showed significant P values in the LR model. This suggests that while some of these variables may not individually differentiate between the analyzed groups, they may interact with other variables in the model to influence the probability to reach a diagnosis. The table below presents the coefficients and P values of these significant variables.

Feature	Coefficient (Importance)	P value
Hepatic failure	2.838622251	.05
Seizure	2.202749367	.03
Generalized hypotonia	1.532873047	.04
Abnormality of the endocrine system	-1.814159083	.04
Abnormal renal morphology	-3.427848427	.01

## Prediction and Analysis of Variants of Uncertain Significance (VUS) Cases

We extended the application of our LR model to include the 15 cases categorized VUS. While these cases were not part of the original model development due to the inherent uncertainty of their diagnostic status, the model presents a useful tool to infer their potential classification.

We applied the predictive model to these cases, allowing us to extrapolate the probable diagnostic status based on the variable patterns observed in the rest of our data. The table below shows the predicted probabilities and diagnostic status for the 15 VUS cases.

ID	Probability	Prediction
(1)	0.64427152	1
(2)	0.78408059	1
(3)	0.29477215	0
(4)	0.65707517	1
(5)	0.20945773	0
(6)	0.23366182	0
(7)	0.55997971	1
(8)	0.69137824	1
(9)	0.67898818	1
(10)	0.45790867	0
(11)	0.88994771	1
(12)	0.17605437	0
(13)	0.94263791	1
(14)	0.06155092	0

The prediction value indicates the predicted diagnostic status, with 1 representing a diagnosis and 0 representing no diagnosis. These predictions provide a potential direction for further investigation of these VUS cases. However, it is essential to note that these are only predictions and should be interpreted cautiously.

eFigure 1. Neonatal Age at Referral for rtGS
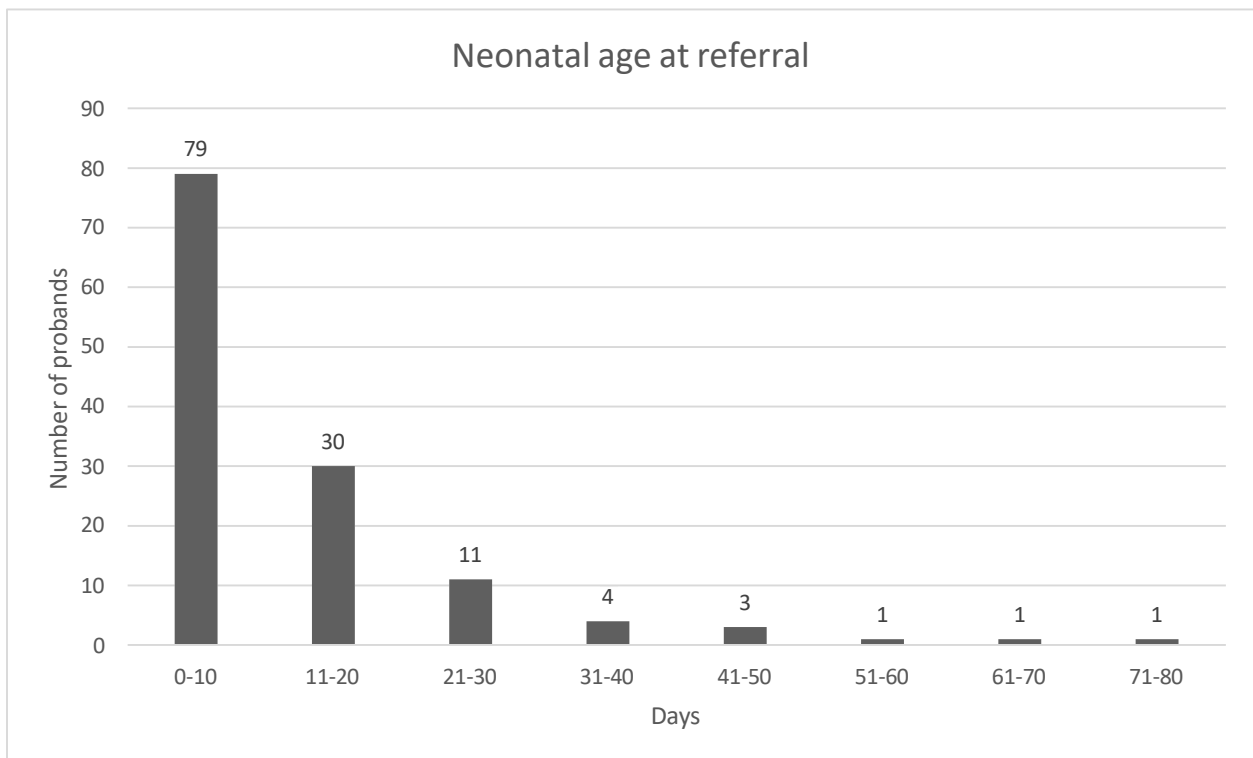
The horizontal axis represents a 10-day increment in age of enrolled probands (0-90 days). The y-axis represents the number of probands (n=130). As shown, 60% (79/130) were ≤10 days of age at referral to rtGS; 83% (109/130) were ≤ 20 days of age. None of the probands were above the 90-day age limit for inclusion.

eFigure 2. Ethnic Backgrounds of Baby Bambi Pilot Population

A. All cohort (n=130). B. Jewish participants (n=68). B. Non-Jewish participants (n=58). D. Inheritance patterns of monogenic disorders diagnosed in Jewish and non-Jewish patients. Fifty-two patients were diagnosed with monogenic disorders (51 with disease-causing SNV and one with myotonic dystrophy).
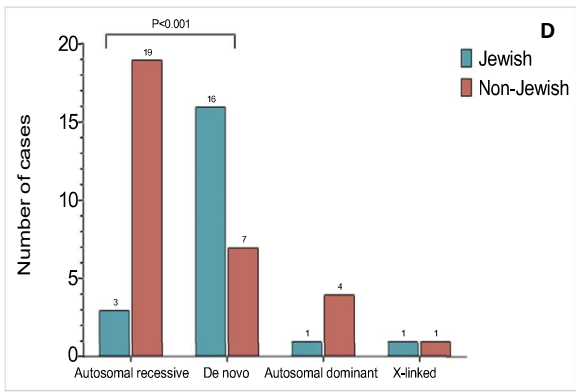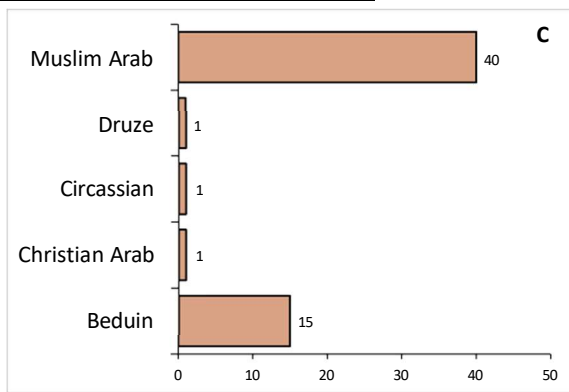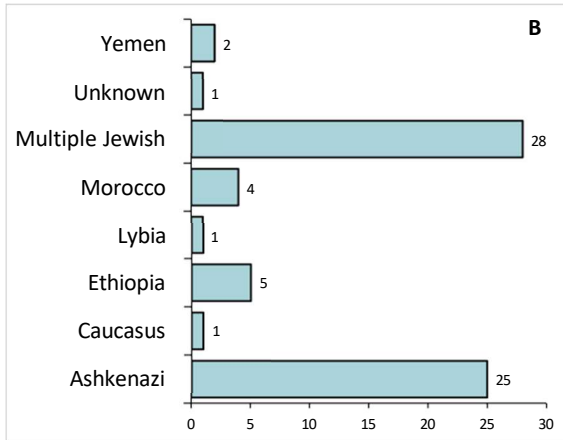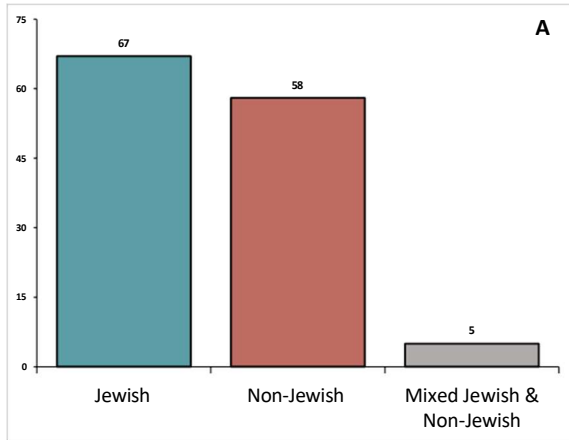
eTable 3. Distribution of Secondary Category Inclusion Criteria (HPO Terms) Among Diagnosed, Possibly Diagnosed, and Undiagnosed Patients

Criterion number		Total	Diagnosed[1]	Possible diagnosis[2]	Undiagnosed[3]	P value[4]
	n^5	130	66	14	50	
1	Abnormality of prenatal development or birth (HP:0001197)	36	18	3	15	.79
2	Abnormal heart morphology (HP:0001627)	30	17	2	11	.96
3	Cardiac conduction abnormality (HP:0031546)	1	1	0	0	NA
4	Abnormality of the skeletal system (includes arthrogryposis) (HP:0000924)	26	17	3	6	.11
5	Abnormality of the skin (HP:0000951)	6	3	1	2	.72
6	Abnormality of metabolism/homeostasis (HP:0001939)	13	5	1	7	.59
7	Abnormality of the endocrine system (includes ambiguous genitalia) (HP:0000818)	18	4	4	1	.02
8	Abnormality of the gastrointestinal tract (HP:0011024)	22	7	2	13	.1
9	Abnormality of blood and blood-forming tissues (HP:0001871)	7	3	0	4	.54
10	Hepatic failure (HP:0001399)	5	3	0	2	1
11	Brain imaging abnormality (HP:0410263)	34	19	2	13	.51
12	Seizure (HP:0001250)	17	11	3	3	.07
13	Generalized hypotonia (HP:0001290)	22	16	1	7	.07
14	Encephalopathy (HP:0001298)	19	10	1	8	.78
15	Abnormal cerebral white matter morphology (HP:0002500)	4	1	0	3	.45
16	Abnormal renal morphology (HP:0012210)	9	1	1	7	.03
17	Renal insufficiency (HP:0000083)	11	4	1	6	.39
18	Abnormality of the respiratory system (HP:0002086)	15	5	1	10	.22
Total number of secondary category inclusion criteria, mean±SD (range)		2.34±1.28 (0-6)	2.27±1.17 (0-5)	1.79±0.77 (1-3)	2.58±1.46 (1-6)	Kruskal-Wallis .28
Additional number of HPO terms, mean±SD (range)		4.33±2.28 (0-12)	4.49±2.37 (1-12)	4.28±2.34 (1-9)	4.14±2.12 (0-8)	Kruskal-Wallis .8

[1]Pathogenic/likely pathogenic variants; [2]variants of unknown significance in gene highly suspected to cause or contribute to phenotype; [3]no disease-causing variants detected; [4]Fisher's exact test; [5] please note, number in each column indicates number of patients with criterion. Patients could be assigned multiple secondary category inclusion criteria, thus are listed in multiple rows. Abbreviations: NA, not applicable; SD, standard deviation.

eFigure 3. Variables of Diagnostic vs Negative rtGS Results
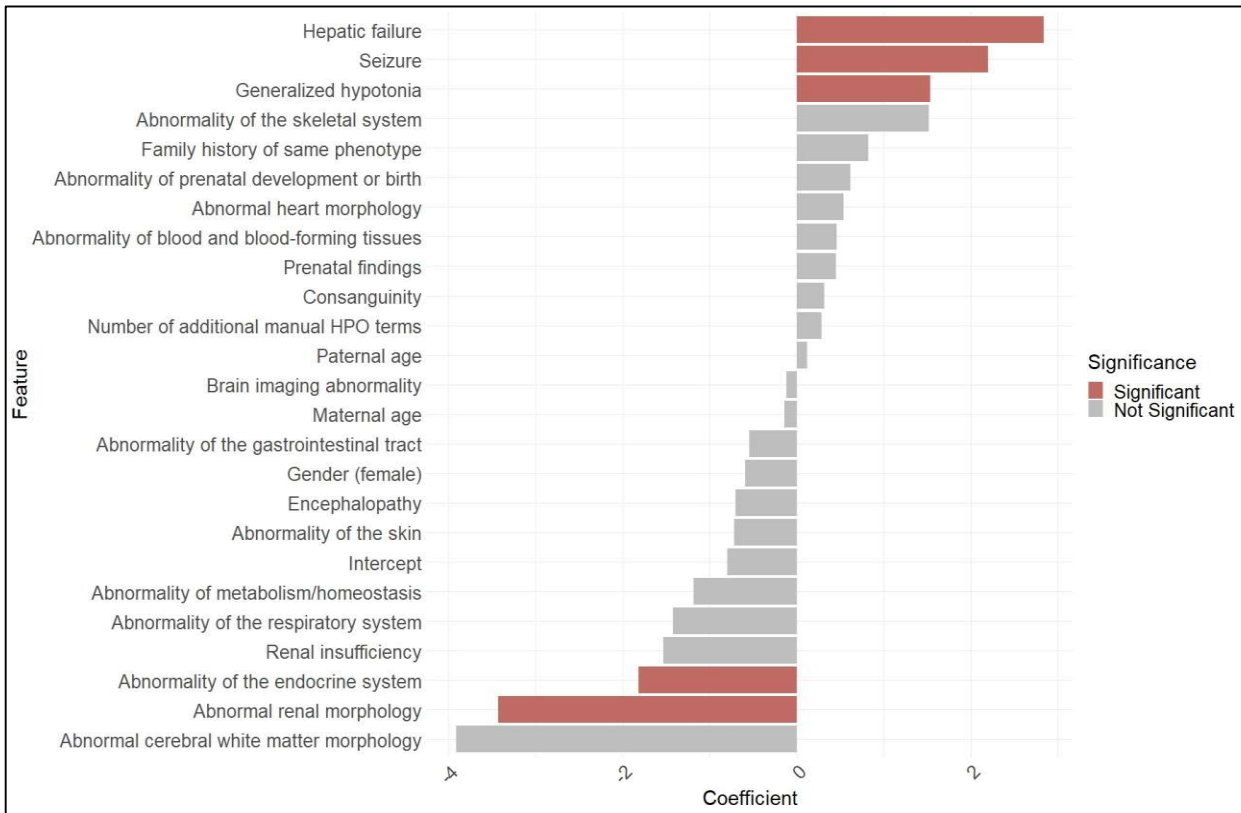
This chart displays the significance and coefficients of clinical variables and rtGS results in the LR model. The brown bars indicate a significant variable, while the grey bars represent a nonsignificant variable. A positive score suggests a positive correlation with a successful rtGS diagnosis, while a negative score indicates a negative correlation. Key findings include significant positive correlations with GS outcomes for the secondary category inclusion criteria: hepatic failure (HP:0001399), generalized hypotonia (HP:0001290), and seizure (HP:0001250). In contrast, renal phenotypes such as abnormal renal morphology (HP:0012210) and abnormality of the endocrine system (HP:0000818) demonstrated a significant negative correlation with GS outcomes.