

A User Manual to Correlagen's Result Reports

Introduction

A gene test is only as useful as its result report. For this reason, Correlagen's result reports provide clear result interpretations in addition to the list of sequence variants detected. Use of the rules-based RightScore™ and RightReport™ systems to generate result reports guarantees consistency of result interpretations across patients and over time.

This document is meant as a "User Manual" to Correlagen's result reports, designed to help the reader find the desired information quickly and understand the data and rules underlying the interpretations fully.

Topics covered here

- Which report sections contain patient-specific information?
- What is the result interpretation based on?
- What does the variant score mean?
- Does the variant score reflect the severity of the test phenotype?
- What is the difference between the variant score and the result interpretation?
- How are variant scores determined?
- How can family testing help to resolve uncertainties around "variants of unknown significance" and "possible disease variants"?
- How are variants named?
- Why is the mRNA isoform (NM number) important?
- What are the possible effects of sequence variants?
- Why does the technical-results table display references?
- What are the technical limitations of the DNA-sequence analysis methodology?
- How can large deletions in a gene be detected?
- How does a family testing report differ from a result report on the index patient?
- How does a "supplementary report after parent testing" differ from a family testing report?
- What are revisable reports?

Which report sections contain patient-specific information?

The **Interpretation section** contains the key message. Interpretations fall into the following five categories:

- associated with the test phenotype
- probably associated with the test phenotype
- possibly associated with the test phenotype
- unknown if associated with the test phenotype
- unlikely to be associated with the test phenotype
- very unlikely to be associated with the test phenotype
- reference sequence

The **Notations or Please Note section** contains qualifying information on the interpretation, eg:

- The test subject is a carrier of a recessive disease. (The interpretation in this case would read “unlikely to be associated with the test phenotype,” since a carrier is typically not affected.)

The **Recommendations section** suggests actions that may (further) clarify the interpretation, eg:

- Parent testing, to clarify if two heterozygous recessive disease variants were both inherited from one parent – in which case the test subject would be a carrier – or separately from both parents – in which case the test subject would be affected.

The **Comments section** elaborates on the Interpretation.

The **Technical Results section** gives the

- state of completion of the sequence analysis. just above the technical-results table.
- sequence variants detected and their interpretation, in the technical-results table.
- identifying information for the reference sequence, in the footnote to the technical-results table

What is the result interpretation based on?

The interpretation is based on applying the rules of Mendelian genetics to the two “most significant” variants found in the tested gene. Two variants are considered, since an autosomal recessive disease may be caused by a combination of two different heterozygous variants. Significance of a variant is determined by the probability of its association with the test phenotype. Eg, a variant scored as associated is more significant than a variant scored as possibly associated, which, in turn, is more significant than a variant scored as unlikely to be associated.

What does the variant score mean?

The variant score is a measure of the probability that a variant by itself can cause the test phenotype. The variant score does not reflect the probability that a variant may “weakly” contribute to the test phenotype or another disease phenotype together with other variants in the same or in different genes. In other words, a variant classified as “not associated with the test phenotype” may or may not constitute a mild risk factor for the test phenotype or another disease phenotype. For this reason, Correlagen lists all variants detected in the result report, regardless of their score.

Does the variant score reflect the severity of the test phenotype?

The variant score given in the report does **not** reflect severity of disease, but only probability of association with disease. In other words, a variant definitively known to be associated with the test phenotype may cause a very mild form of the disease, while a variant scored as possibly associated with the test phenotype may cause a particularly severe form of the disease.

How does the variant score relate to the result interpretation?

The variant score reflects the relationship of an individual variant to the test phenotype. The result interpretation considers the variant scores of the two most significant variants in the context of patient-specific parameters, such as the variant zygosity and patient sex.

How are variant scores determined?

Variant scores are based on data drawn from publications, publicly available databases, and/or Correlagen's own sequencing data. Importantly, all data, including all publication data, are evaluated using the RightScore™ algorithm. This guarantees consistent variant scoring and allows pooling of data from many different sources to derive the variant score. However, this also means that the variant score given in the report may differ from the variant score proposed by the authors of a publication. It should also be noted that published data cannot be guaranteed to be correct. The variant score is assembled from up to three component scores:

The predicted functional change (pFXN) score:

The pFXN score reflects a theoretical prediction about how a variant in the genomic DNA will affect the synthesis and/or function of the encoded protein. Both the nature of the change caused in the gene product and the location of the change in the gene product are taken into account. The effect of some variants is relatively easy to predict. For example, truncation of a protein due to a nonsense mutation is very likely to cause loss-of-function of that protein. The effect of a missense mutation, in contrast, is much more difficult to predict, since it depends both on the difference between the old and the new amino acid and the location of the affected amino acid in the protein. RightScore™ currently uses the BLOSUM50 matrix to assign a pFXN score to missense mutations (BLOSUM = Blocks Substitution Matrix, <http://blocks.fhcrc.org/index.html>). Creation of a new splice site by a nucleotide substitution is especially difficult to predict.

The genotype-phenotype correlation (G/P) score:

The G/P score is derived from *in-vivo* data about the association – or the lack of association – of a variant with a specific phenotype. If a variant is observed more frequently in the general (healthy) population than would be expected for a pathogenic variant, given the disease prevalence and the mode of inheritance, then this variant is assumed to be non-pathogenic. If a variant is observed only in diseased individuals and not in the general (healthy) population, it is assumed to be pathogenic. The exact G/P score depends on such parameters as inheritance mode and prevalence of the test phenotype, variant frequency in the general (healthy) population, number of diseased individuals with the variant, and consistency of co-migration of variant and disease within families.

The actual functional change (aFXN) score:

The aFXN score reflects the effect of a variant on the synthesis and/or function of the encoded protein in an experimental system. While experimental systems can provide powerful information, the results must also be interpreted with caution, since an experimental environment lacks many of the complexities of the actual *in-vivo* environment.

How can family testing inform the variant score?

If all affected family members harbor the variant in question, this variant is likely to be associated with the test phenotype. If some affected family members do not harbor the variant in question, it is less likely to be associated with the test phenotype. Clarifying the significance of a variant through family testing not only benefits this particular family, but also other families who harbor the same mutation.

How are variants named?

Variants are described by the type of change and its location within or relative to the gene. In the published literature, several different conventions are used for describing sequence variants. Therefore, the same variant may be described differently in different publications. Correlagen numbers and names all variants according to the system suggested by the Human Genome Variation Society, regardless of the convention used in referenced publications. More information on these systems can be found at <http://www.genomic.unimelb.edu.au/mdi/mutnomen> and at <http://www.correlagen.com/reportinfo/syntax.jsp>.

Why is the mRNA isoform (NM number) important?

The DNA sequence for a gene often contains coding segments (exons) as well as interspersed non-coding segments (introns). An mRNA isoform is defined by the permutation of exons contained in it. Since the mRNA isoform determines if a sequence variant is considered as exonic or intronic, it may impact interpretation of a variant's effect on the encoded protein.

What are the possible effects of sequence variants?

Changes in the gene sequence may have a deleterious effect (1) on the synthesis or processing of the mRNA transcribed from the DNA and/or (2) the protein translated from the mature mRNA. Here are some examples:

Splice-site mutations: A sequence variant can destroy an existing splice site at an exon/intron or intron/exon border or create a new splice site in the middle of an exon or an intron. Both types of variations can lead to altered mRNA processing and a dramatically different mature mRNA sequence, which translates into a dramatically different protein sequence.

Change in mRNA stability: A sequence variant can lead to reduced mRNA stability, which translates into lower amounts of translated protein.

Missense mutations: A sequence variant can lead to replacement of one amino acid by another in the protein. Such missense mutations are commonly caused by a single-nucleotide substitution, as shown in the example below:

<i>Before mutation</i>														
G	G	G	C	T	T	A	A	A	A	C	A	G	C	G
Glycine			Leucine		Lysine			Threonine		Alanine				
<i>After mutation</i>														
G	G	G	C	C	T	A	A	A	A	C	A	G	C	C
Glycine			Proline		Lysine			Threonine		Alanine				

Nonsense mutations: A sequence variant can introduce a stop codon in the middle of the coding region, leading to truncation of the protein. Such nonsense mutations are commonly caused by a single-nucleotide substitution, as shown in the example below:

<i>Before mutation</i>														
G	G	G	T	T	G	A	A	A	A	C	A	G	C	G
Glycine			Leucine		Lysine			Threonine			Alanine			
<i>After mutation</i>														
G	G	G	T	A	G	A	A	A	A	C	A	G	C	C
Glycine			stop											

Frameshift mutations: A sequence variant can give rise to a shift in the reading frame, leading to a complete change of the amino acid sequence downstream of the frameshift site and, since stop codons tend to be enriched in the two unused reading frames, often to a truncation of the protein. A frameshift mutation is caused by a net deletion or net insertion of a number of nucleotides not divisible by 3. Of note, the amino acid sequence may not change until several amino acids downstream of the actual frameshift site, as shown in the example below:

<i>Before mutation</i>														
G	G	G	C	T	T	A	A	A	A	C	A	G	C	G
Glycine			Leucine		Lysine			Threonine			Alanine			
<i>After mutation</i>														
G	G		C	T	T	A	A	A	A	C	A	G	C	C
Glycine			Leucine		Lysine			Glutamine			Arg...			

In-frame deletions and/or insertions: A net deletion and/or insertion of a number of nucleotides divisible by 3 typically leads to deletion and/or insertion of one or more amino acids from/into the protein. Since the reading frame is preserved, the amino acid sequence downstream of the insertion and/or deletion site does not change. Depending on the inserted sequence and the site of the insertion and/or deletion relative to the reading frame, a missense mutation may result, as shown in the example below:

<i>Before mutation</i>														
G	G	G	C	T	T	A	A	A	A	C	A	G	C	G
Glycine			Leucine		Lysine			Threonine			Alanine			
<i>After mutation</i>														
G	G	G	C	G	C	A	T	T	A	A	A	A	C	A
Glycine			Arginine		Isoleucine			Lysine			Threonine			

Why does the technical results table display references?

References for variants in the technical-results table contain data relevant to determining the significance of these variants for the test phenotype. It is important to point out that Correlagen's reports are based on the GeneExplorer™ database of variants and linked publications, which is highly curated and updated on a weekly basis. Correlagen encourages readers of the result report to obtain and read the referenced publications. Of note, the numbering and naming of variants used in the report may differ from the convention used in a publication, and the report's interpretation of a variant's significance for the test phenotype may differ from the authors' interpretation.

What are the technical limitations of the DNA sequencing method?

Correlagen's sequence analysis is based on PCR amplification of the target DNA sequence, followed by dideoxy sequencing of both DNA strands of each PCR product. For genes that are present on two chromosome copies (i.e., autosomal genes and X-linked genes in females), both chromosomal gene copies serve as templates for PCR amplification. Sequencing traces obtained from the PCR products therefore reflect a mixture of the gene sequences present on these two chromosome copies. If the same nucleotide is present at a given position in the gene sequence on both chromosome copies (homozygosity), a single signal corresponding to that nucleotide will appear in the sequencing traces. If different nucleotides are present at a given position in the gene sequence on the two chromosome copies (heterozygosity), two overlapping signals corresponding to the two nucleotides will appear in the sequencing traces. Information about which particular chromosome copy a sequence variant is located on is lost during sequence analysis. Therefore,

- The sequencing results do not allow any conclusion as to whether two different heterozygous sequence variants are present on the same or on different chromosome copies. This limitation can be overcome by parent testing, since one chromosome (and the variants on it) is inherited from the father and the other from the mother.
- The sequencing results cannot distinguish if both or only one of the two chromosome copies served as a template for PCR amplification. If only one of the two chromosome copies served as a template, variants on the other chromosome copy would be missed. For example, a large deletion in or of the gene on one chromosome copy would constitute a pathogenic variant. The gene region would only be amplified from the other chromosome copy, and the deletion would not be directly detected. In this case, however, all variants present on the amplified chromosome copy would appear to be homozygous, so that uniform homozygosity of many sequence variants within a gene can serve as an indication of a large heterozygous gene deletion. Of note, such large gene deletions, also known as copy-number variations, appear to be more common than originally assumed.

How can large deletions in a gene be detected?

Large gene deletions can be detected by such methods as comparative genomic hybridization, real-time quantitative PCR, and/or gap PCR. Correlagen currently uses gap PCR to detect certain large deletions in defined genes. PCR primers are selected such that they border the deletion site on either side. The size of the PCR product generated from these primers depends on whether the deletion has occurred or not.

What is different about a Family-Testing Report?

Interpretations given in a family testing report are based on the assumption that the tested gene region contains the true familial mutation, i.e., the variant responsible for the test phenotype in this family. Absence of the familial mutation shows that the family member is not at any increased risk of being affected with or developing the test phenotype, even if large parts of the gene sequence were not determined. In the index patient, in contrast, absence of pathogenic variants after partial sequence analysis would lead to an interpretation of "unknown if associated." Importantly, absence of the familial mutation can only rule out the test phenotype in family members of the index patient if the familial mutation is scored as associated or probably associated with the test phenotype. Sometimes, a variant assessed as possibly associated with disease is suspected to be the familial mutation. Absence of such a possible disease variant cannot rule out the test phenotype in family members of the index patient, since it is not certain that this variant is the cause of test phenotype in this family.

What is different about a Supplementary Report after Parent Testing?

The supplementary report after parent testing confirms or modifies the original report on the child's genotype. In other words, the "parent testing report" is issued for the child's benefit. In contrast, the family testing report is issued for the benefit of the family member tested.

What are revisable reports?

Correlagen's services to the patient and the ordering physician do not end with the issuance of a result report. Instead, Correlagen is committed to an ongoing relationship that involves updates to the ordering physician and/or the patient as relevant new information on a variant, gene, or test phenotype emerges. For example, a revised result report may be issued when the score for a variant found in a patient changes significantly and this change has implications for the report interpretation. We may also ask for additional clinical information on the patient to refine variant and result interpretation, or we may offer sequencing for other genes that have been newly associated with the same test phenotype. Recognizing the rapid evolution of knowledge in the field of genomic medicine, Correlagen's result reports are intended as "living documents" that can evolve with advances in scientific knowledge.