

### More information on Correlagen's RightReport™

Below are answers to common questions about our result reports. Please let us know if you have other report-related questions. We value your feedback.

[How does Correlagen number and name variants?](#)

[Why does Correlagen specify the mRNA isoform \(NM number\)?](#)

[What are possible effects of sequence variants \(types of mutations\)?](#)

[What do Correlagen's variant scores mean?](#)

[How are Correlagen's variant scores determined?](#)

[How can family testing help to improve variant scoring?](#)

[Are there other systems for scoring variants?](#)

[Why does Correlagen report every variant, including polymorphisms?](#)

[Why does Correlagen link references to variants in the technical results table?](#)

[What are the technical limitations of the DNA sequence analysis that Correlagen performs?](#)

[What methodology does Correlagen use to detect large deletions?](#)

[How can parent testing help to interpret the sequencing results for a child?](#)

[What if no sequence variants are detected in a patient with clear clinical symptoms of the test phenotype?](#)

### How does Correlagen number and name variants?

In the published literature, several different conventions are used for numbering and naming sequence variants. In other words, the same sequence variant may be numbered and named differently in different publications. Correlagen numbers and names all variants according to the system suggested by the Human Genome Variation Society (<http://www.genomic.unimelb.edu.au/mdi/mutnomen>), regardless of the numbering and naming conventions used in the publications describing the variants.

- **Which nucleotide is +1?**

In the published literature, authors may start counting from the transcription start site, i.e., the beginning of the mRNA sequence, or from the beginning of the cDNA they generated, which may or may not reflect the beginning of the mRNA. Correlagen designates the start of the coding sequence of a gene as +1.

- **Are intronic sequences counted?**

Many genes have non-coding regions (introns) interspersed between coding regions (exons). Correlagen counts only nucleotides in exonic regions. Nucleotides in intronic regions or in regions upstream or downstream of the coding regions are numbered relative to exonic regions.

Please [click here](#) for a detailed description of the numbering and syntax rules used by Correlagen, as well as for specific examples.

### Why does Correlagen specify the mRNA isoform (NM number)?

Often, the same gene sequence can give rise to several mRNA isoforms. For genes with many exons, different mRNA isoforms may contain sequence from different permutations of exons. The exons reflected in a particular mRNA isoform define the actual coding region and thus numbering of the sequence variants. By determining if a sequence variant is considered as exonic or intronic, the mRNA isoform also may impact interpretation of a variant's effect on the encoded protein.

### What are possible effects of sequence variants (types of mutations)?

Sequence variants, or mutations, are classified both according to the change they cause in the gene sequence, i.e., the DNA sequence, and according to the effect they have on synthesis or processing of the mRNA transcribed from the DNA and/or the protein translated from the mature mRNA. The following is a brief description of the most common types of mutations. It is not a complete listing of all types of possible mutations.

- On the DNA level, the most common classes of mutations are:
  - Substitutions of one nucleotide for another nucleotide, changing the nucleotide sequence but not the nucleotide number in the gene sequence.
  - Substitutions of a group of nucleotides for a group of different nucleotides. These mutations are commonly referred to as “indels,” or insertions-deletions, and can potentially lead to both a change in nucleotide sequence and a change in nucleotide number in the gene sequence.
  - Deletions of one or more nucleotides
  - Insertions of one or more nucleotides
- On the mRNA level, common deleterious effects of sequence variants include:
  - Alteration of splicing: 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.
- On the protein level, common effects of sequence variants include:
  - A change of one amino acid in the protein into another. Such missense mutations are commonly caused by a single-nucleotide substitution, as shown in the example below:

G G G	C T T	A A A	A C A	G C G
Glycine	Leucine	Lysine	Threonine	Alanine
G G G	C C T	A A A	A C A	G C G
Glycine	Proline	Lysine	Threonine	Alanine

- An introduction of 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:

G G G	T T G	A A A	A C A	G C G
Glycine	Leucine	Lysine	Threonine	Alanine
G G G	T A G	A A A	A C A	G C G
Glycine	stop			

- 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:

G G G	C T T	A A A	A C A	G C G
Glycine	Leucine	Lysine	Threonine	Alanine
G G	C T T A	A A A	C A G	C G
Glycine	Leucine	Lysine	Glutamine	Arg...

- An in-frame deletion and/or insertion of one or more amino acids from/into the protein that does not alter the amino acid sequence downstream of the insertion and/or deletion site, as shown in the example of a 3-nucleotide insertion (TCA) below:

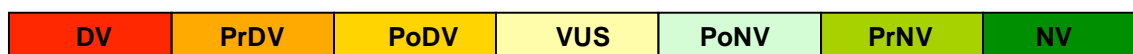
G G G Glycine	C T T Leucine	A A A Lysine	A C A Threonine	G C G Alanine
G G G Glycine	T C A Serine	C T T Leucine	A A A Lysine	A C A Threonine

### What do Correlagen's variant scores mean?

Correlagen's variant scores, shown in the column "Relationship to *disease phenotype*" in the technical results table of the report, are a measure of the probability that a particular variant by itself can cause a defined monogenic disease phenotype. The variant score does not reflect the probability that a variant may be associated with the test phenotype – or another disease phenotype – in an oligogenic or polygenic rather than monogenic manner. In other words, a variant score of "(probably/possibly) not associated" does not exclude the possibility that a variant may "weakly" contribute to a disease phenotype in association with several other variants in the same or different genes. The score also does not reflect the severity of any associated monogenic disease phenotype.

There are seven score categories:

Score Category	Abbreviation	Is the variant by itself able to cause the defined disease phenotype?
Disease variant	DV	Yes
Probable disease variant	PrDV	Probably
Possible disease variant	PoDV	Possibly
Variant of unknown significance	VUS	Unknown
Possible normal variant	PoNV	Unlikely
Probable normal variant	PrNV	Very unlikely
Normal variant	NV	No



high ←———— probability of association with monogenic disease phenotype —————→ low

### How are Correlagen's variant scores determined?

Variant scores are assembled from several 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 (e.g., truncation of the gene product due a nonsense mutation, alteration in the amino acid sequence of the gene product due to a missense mutation, or deletion of one exon due to a splice-site mutation) and the location of the change in the gene product (i.e., the evolutionary conservation of the affected region) are taken into account. The effect of some variants is relatively easy to predict; truncation of a protein due to a nonsense mutation, for example, 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. 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 reflects observed *in-vivo* data about the association – or the lack of association – of a variant with a specific phenotype. There are different types of data that can be used to determine a G/P score.
    - Frequency of a variant in the general population: If a variant is significantly more common in the general 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 normal. Note that frequency data can only be used to assign a positive G/P score. In other words, a variant is never assumed to be pathogenic just because it is rare in the general population.
    - Frequency of a variant in the diseased population: The odds that a variant and a disease phenotype are associated are calculated as the ratio of the probability of the observed co-occurrence of a variant and the disease phenotype assuming association over the probability of the observed co-occurrence of a variant and the disease phenotype assuming no association. Different algorithms are used depending on whether the data are derived from related or from unrelated individuals.
  - **The actual functional change (aFXN) score:**  
The aFXN score reflects an observed *in-vitro* effect of a variant on the synthesis and/or function of the encoded protein. While *in vitro* experiments can provide powerful information, the results must also be seen with caution, since the *in-vitro* environment lacks many of the complexities of the *in-vivo* environment.

The data used to calculate the component scores can be drawn from several different sources, including publications, publicly available databases, or Correlagen's own sequencing data. Since all data, including publication data, are evaluated using Correlagen's scoring algorithms, Correlagen's variant score may differ from the variant score proposed by the authors of a publication. It should also be noted that Correlagen has no guarantee, beyond peer-review of the papers before publication, that published data are correct.

### How can family testing help to improve variant scoring?

If Correlagen obtains sequence information and clinical information for the blood relatives of a patient, these genotype phenotype-correlation data can be included in the variant analysis and, possibly, improve the level of certainty that a familial variant is or is not associated with the test phenotype.

### Are there other systems for scoring variants?

There currently is no standardized system for scoring variants, although all systems are based on the same basic principles. Some commonly used algorithms for determining the predicted functional score of missense mutations include the BLOSUM matrices (BLOSUM = Blocks Substitution Matrix, <http://blocks.fhcr.org/index.html>), SIFT (Sorting Intolerant From Tolerant, <http://blocks.fhcr.org/index.html>), and PolyPhen (Polymorphism Phenotyping, <http://genetics.bwh.harvard.edu/pph/>). Correlagen currently uses the BLOSUM50 matrix to assign a pFXN score to missense mutations.

### Why does Correlagen report every variant, including polymorphisms?

Correlagen defines "sequence variant" as any change from the reference sequence, regardless of the variant score. If a variant is classified as "not associated with the test phenotype," it is

considered a benign polymorphism. The reasons for reporting polymorphisms as well as pathogenic sequence variants are:

- Correlagen considers it appropriate to give the physician and the patient all of the sequencing results, not just selected portions.
- A variant classified as probably not associated, i.e., a probable benign polymorphism, may (turn out to) contribute to the test phenotype or another disease phenotype in a multi-variant and, possibly, polygenic fashion. In that case, it would be important for the patient to know if he or she harbored that variant.

### **Why does Correlagen link references to variants in the technical results table?**

References for variants in the technical results table contain data relevant to determining the significance of these variants for the disease phenotype. Correlagen encourages readers of the result report to obtain and read the referenced publications. Of note, Correlagen's numbering and naming of variants may differ from the convention used in a publication, and Correlagen's interpretation of the significance of a variant for the test phenotype may differ from the authors' interpretation.

### **What are the technical limitations of the methodology for DNA-sequence analysis that Correlagen uses?**

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 template 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 about whether two different heterozygous sequence variants are present on the same or on different chromosome copies. This information may be important, since two heterozygous pathogenic sequence variants can cause a recessive disease phenotype only if they are located on different chromosome copies. Of note, this limitation can be overcome through parent testing (see ["How can parent testing help to interpret the sequencing results for a child?"](#)).
- The sequencing results cannot distinguish if both or only one of the two chromosome copies served as a template for the PCR amplification. If only one of the two chromosome copies served as a template, all nucleotides found in that amplicon would simply appear homozygous. Hidden behind such apparent homozygosity could be sequence variants present on the non-amplified chromosome copy. One specific case of such a sequence variant would be a large deletion leading to complete loss of the gene region in question from one chromosome copy. Of note, such copy-number variations appear to be more common than originally assumed. To overcome this limitation, Correlagen is in the process of implementing methodology to detect copy-number variations in genes.

### **What methodology does Correlagen currently use to detect large deletions?**

Correlagen currently can detect certain large deletions in defined genes. PCR primers are selected such that they border the region typically deleted on either side. The size of the PCR product generated from these primers depends on whether the deletion has occurred or not.

### How can parent testing help to interpret the sequencing results for a child?

As discussed under "[What are the technical limitations of the DNA sequence analysis that Correlagen performs?](#)," the sequencing results do not allow any conclusion about whether two heterozygous pathogenic sequence variants are located on the same or on different chromosome copies. This question can be answered by parent testing, since one chromosome copy is inherited from the father and the other from the mother. This information is often important since a recessively inherited disease is only expressed if the variants are located on different chromosome copies and a dominantly inherited disease is often more severe if the variants are located on different chromosome copies. Information about whether two variants are located on the same or on different chromosome copies is also relevant for genetic counseling or in a situation where expression of the disease is influenced by whether a pathogenic variant was inherited from the father or from the mother.

### What if no sequence variants are detected in a patient with clear clinical symptoms of the test phenotype?

If Correlagen does not find a sequence variant in a given gene for a patient who has clear symptoms of the test phenotype, the following possibilities should be considered:

- Disease in the patient is caused by sequence variation in a gene other than the one sequenced. Other gene tests for the same test phenotype should be considered, if available.
- Disease in the patient is due to mosaicism for sequence variation in the gene sequenced. Depending on which cell lineage is affected by mosaicism and the ratio of cells containing and not containing the variation, Correlagen's sequencing methodology may or may not be able to detect the sequence variant. Repeat testing is suggested with DNA derived from another type of sample, e.g., a buccal swab if the original sample was blood.
- Sequence analysis failed to detect the pathogenic sequence variant causing the disease (see "[What are the technical limitations of the DNA sequence analysis Correlagen performs?](#)"). Correlagen will soon add extended-testing options that will include sequence analysis of regulatory regions and methodology to detect copy-number variation.