

## Frequently Asked Questions about Correlagen’s CardioGeneScan (CGS)

### What is the CardioGeneScan?

#### **...comprehensive**

The CGS is the most comprehensive genetic test for cardiac disease that is available today. It tests the sequences of over 100 genes known to be associated with genetic cardiac disease. All major known genetic causes of both isolated and syndromic types of familial cardiomyopathy, arrhythmia, aortopathy, Noonan syndrome and related disorders, congenital heart disease, and coronary artery disease/hypercholesterolemia are included, as well as many minor causes.

#### **...clear**

The CGS combines clear, simple indications with easy-to-read result reports. The broad range of genes included makes the CGS applicable to all types of familial, i.e., genetically-based, cardiac disease. Our specially designed results reports highlight the most relevant findings, based on our interpretation of all related publications and resources. The CGS can be ordered for six different indications: familial cardiomyopathy, familial arrhythmia, familial aortopathy, Noonan syndrome and related disorders, genetically-based congenital heart disease, and familial coronary artery disease/hypercholesterolemia. Only genes relevant to the ordered indication or indications will be analyzed and interpreted in the results report.

#### **...current**

The CGS will be kept current by periodic updating with genes that have been newly shown to be associated with familial cardiac disease. Variant interpretation will also be kept current, and updated results reports will be issued if new literature or accrued data change the meaning of a critical variant or variant combination. Patients will have the opportunity to opt out of allowing re-testing of their sample with updated versions of the CGS or re-interpretation of their sequencing results.

#### **...cost effective**

The CGS is available at affordable price points for any of the six indications, with minimal patient obligation for patients with insurance coverage. Re-testing of patient samples in which no pathogenic variants were previously detected, either for different indications or with updated versions of the CGS, will also be offered.

### What is the clinical sensitivity of the CardioGeneScan?

Panel	Disease Name	Sensitivity
Familial Cardiomyopathy	Hypertrophic Cardiomyopathy	>65%
	Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy	>50%
	Dilated Cardiomyopathy	>40%
	Other Cardiomyopathies (restrictive, left ventricular non-compaction)	>10%
Familial Arrhythmia	Long QT Syndrome/Brugada Syndrome	>75%
	Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy	>50%
	Catecholaminergic Polymorphic Ventricular Tachycardia	>50%
Familial Aortopathy	Marfan Syndrome/Loeys-Dietz Syndrome/vascular-type Ehlers-Danlos Syndrome	>90%
	Thoracic Aortic Aneurysm and Dissections	>15%
Noonan Syndrome and Related Disorders	Noonan Syndrome, LEOPARD Syndrome, Cardiofaciocutaneous Syndrome, Costello Syndrome	>70%
Familial Congenital Heart Disease	Atrial Septal Defect	>12%
	CHARGE Syndrome	>75%
Early-onset Coronary Artery Disease/ Familial Hypercholesterolemia	Familial Hypercholesterolemia	>90%

### What methodology is the CardioGeneScan based on?

The CGS is primarily based on novel single-molecule sequencing on the Helicos® Genetic Analysis system, which allows sequencing of millions of bases in parallel, with high detection sensitivity. In addition, conventional “Sanger” sequencing may be used for parts of the test. All positive findings from single-molecule sequencing are confirmed using conventional “Sanger” sequencing.

Single-molecule sequencing on the Helicos Genetic Analysis System, or the HeliScope™ Sequencer, uses a new type of sequencing methodology that is often referred to as “third-generation” or “massively parallel” DNA sequencing. To prepare DNA for sequencing on the Heliscope, the DNA is randomly broken into small fragments, and a poly(dA) tail is added to the 3’ end of each fragment. Billions of these poly(dA)-tailed DNA fragments are then bound in parallel to a special glass surface (the “flow cell”) coated with oligo(dT)s for capturing the fragments. The bound fragments serve as templates for generating new, complementary DNA strands in a process referred to as sequencing-by-synthesis. In each sequencing cycle, only one specific, fluorescently labeled nucleotide is added to the flow cell. For example, during the 1<sup>st</sup> sequencing cycle, a fluorescently labeled dCTP could be added. This nucleotide will only be incorporated into those growing complementary DNA strands that need a C as the next nucleotide. A laser is used to cause the nucleotide to emit light, and a picture is taken of the flow-cell surface: DNA strands that have incorporated a C will emit light, while DNA strands that have not incorporated a C will appear dark. The sequencing cycle is completed by removal of the fluorescent label, and the entire process is repeated with a different nucleotide (such as dGTP), and then a third and a fourth nucleotide, for a total of 120 cycles. The images are converted real-time into corresponding strings of bases, commonly referred to as “reads,” which recapitulate the 3’ terminal 25 to 60 bases of each fragment. The reads are then compared to the reference sequence for the DNA that was analyzed (eg, the human genome sequence if human genomic DNA was used). Since any given string of 25 bases typically only occurs once in the human genome, most reads can be “aligned” to one specific place in the human genome. Finally, a consensus sequence of each genomic region is built from the available reads and compared to the exact sequence of the reference at that position. The probability that the consensus sequence is correct at any particular position in the genome largely depends on the number of reads covering this position (known as “coverage”). Any differences between the consensus sequence and the reference are called as sequence variants. More detailed information about the process can be found on the manufacturer’s website.

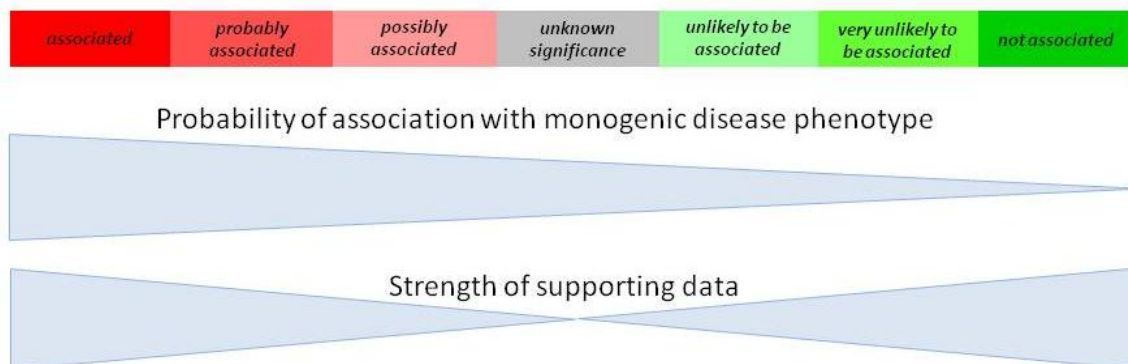
Since only certain genes of interest will be sequenced for the CGS, these genes first have to be separated from the rest of the genomic DNA. Correlagen uses two methods for this: The desired parts of the genome are either physically “fished out” by binding to specific complementary nucleic acid probes, or, alternatively, amplified relative to all other parts of the genome by a polymerase chain reaction (PCR) with specific primers. Once the target genes have been isolated, they can be fragmented and sequenced as described above.

### What is the analytical sensitivity of the CardioGeneScan?

The CGS currently detects 100% of missense and nonsense mutations and 90-95% of all small deletions and insertions. In addition, large deletions spanning parts or all of a gene are expected to become detectable soon, increasing the sensitivity of the CGS over that achievable by conventional Sanger sequencing alone.

## How does Correlagen indicate variant significance?

Correlagen assigns a score to each variant that reflects the probability that a variant is associated with monogenic disease. This probability can range from very high (“associated with monogenic disease”) to very low (“not associated with monogenic disease”). Importantly, the score also expresses the strength of the supporting data, or the confidence that the score is correct. A score of “associated” or “not associated” is well supported by data. In contrast, a score of “possibly associated” or “unlikely to be associated” carries a high degree of uncertainty and cannot by itself support diagnosis of disease. The figure below illustrates this concept.



## How does Correlagen determine the clinical significance of a variant?

Correlagen’s variant scores are based on the following considerations:

**Has the variant been observed in the general population or normal controls?** If a variant is observed more frequently in the general population than is compatible with the prevalence and mode of inheritance of the disease, then this variant is assumed to be non-pathogenic. Data for variant frequency in the general population are derived from dbSNP (NCBI EntrezGene), from peer-reviewed publications, and/or from prevalence studies conducted at Correlagen.

**Has the variant been observed in affected individuals?** If a variant is observed only in diseased individuals and not in the general (healthy) population, it is assumed to be at least possibly associated with disease. The probability of association depends on such parameters as the number of diseased individuals with the variant and the consistency of co-occurrence of variant and disease within families. The number of variant occurrences in affected individuals is derived from the peer-reviewed published literature.

**What effect does the variant have in a controlled experimental system?** A significant effect of a variant on the synthesis, cellular location, and/or function of the encoded protein in an experimental system suggests that the variant is pathogenic. While experimental systems can provide powerful information, the results must also be seen with caution, since an experimental environment lacks many of the complexities of the actual *in-vivo* environment. Data for variant effect in an experimental system are derived from the peer-reviewed published literature.

**What is the predicted effect of the variant on synthesis and/or function of the encoded protein<sup>3</sup>?** If the variant leads to truncation of the gene product due a nonsense mutation or a frameshift mutation, it is assumed to be pathogenic (for diseases related to loss-of-function mutations). If the variant affects one of the highly conserved donor or acceptor splice sites, it is predicted to lead to exon skipping and is assumed to be pathogenic. If the variant leads to a missense variant, it is considered to be possibly pathogenic, in absence of other information. If the variant is located in the coding region away from exon/intron junctions and does not lead to a change in the amino acid sequence (synonymous variant), it is considered

unlikely to be pathogenic, although pathogenicity cannot be excluded. If a variant is located in the coding sequence close to an exon/intron junction or in an intron away from the exon/intron junction, its effect cannot be predicted, and it is classified as a variant of unknown significance.

**Several prediction algorithms** (eg, SIFT, PolyPhen, Align-GVGD)<sup>1</sup> have been developed in an attempt to predict the impact of a missense variant on protein function. Prediction algorithms are typically based on evolutionary conservation, structure of the protein at the site of the variant, and/or amino acid properties. While Correlagen routinely uses such algorithms for variant evaluation, it does not base the variant score on a prediction from any single one of the algorithms, since their specificity and sensitivity are limited and their predictions frequently contradict each other. (<sup>1</sup> <http://blocks.fhcrc.org/sift/SIFT.html>, <http://coot.embl.de/PolyPhen/>, [http://agvgd.iarc.fr/agvgd\\_input.php](http://agvgd.iarc.fr/agvgd_input.php))

**Additional considerations** for scoring include, but are not limited to, co-occurrence of a variant with known pathogenic variants, occurrence of a variant in mutually exclusive disease phenotypes (for example, very high and very low blood sugar), predicted effect of synonymous variants on splicing, and certain gene-specific and/or disease-specific properties.

**Conflicts between Correlagen's variant score and the variant score given in a publication** can occur. We determine the variant score based on a combined analysis of all data we could find. In other words, we use the data given in a publication, not the author's conclusion. Occasionally, Correlagen's variant score may therefore differ from the author's interpretation.

#### **Correlagen's variant scoring is evidence based and transparent.**

We realize that some clients would like to form their own opinion about variant significance. As a service to these clients, we provide a "variant scoring sheet" upon request. This scoring sheet lists all the information we could find for a variant at the time of scoring. All information is well referenced and sorted by type of information, eg, variant prevalence in normals, variant occurrences in diseased, functional effect measured *in vitro*, or the results of various prediction algorithms. If no data are listed in a section, we couldn't find any - but we always search! The scoring sheet also contains annotations from Correlagen's variant scientists, such as short descriptions of an *in-vitro* experimental system used in a publication or comments on specific pedigree data.

#### **What if no pathogenic variants are found?**

If no pathogenic variants are found under the ordered indication, reflexing to any of the other indications covered by the CGS is available. In addition, the sample can be re-tested at a reduced price when new, expanded versions of the CGS are released. Patients will have the opportunity to opt out of allowing re-testing of their sample for other indications or with updated versions of the CGS.

#### **What if only variants of unknown significance are detected?**

Correlagen will issue updated result interpretations to the ordering physician as it becomes aware of newly available information that clarifies the clinical significance of a variant. This service is currently free.

#### **What if variants associated with non-cardiac manifestations are found?**

Some of the genes included in the CGS may be associated with non-cardiac manifestations, such as neurological symptoms. Patients will have the opportunity to opt out of having a variant included in the results report that is known to be associated solely with non-cardiac manifestations. For genes associated with syndromic forms of cardiac disease, the same variant can be related to cardiac and non-cardiac manifestations. Such variants will be included in the clinical results report.

### **Will normal variants be reported?**

Variants that show a high prevalence in the general population are classified as normal. Correlagen uses data from the publicly funded 1000-Genomes Project and from its own prevalence studies to identify such normal variants. Normal variants will be listed in a summary table at the end of the results report.

### **When and why does Correlagen issue partial reports?**

Big tests sometimes take a while to complete. If we find a pathogenic variant or variant combination partway through an analysis that supports a diagnosis of disease, we may send out a partial report, to inform our clients and their patients as soon as possible of this significant finding. Once the analysis is complete, we will issue a final report, unless this requirement is waived by the ordering physician. We may also send out a partial report if 90% or more of a test has been completed, and we are experiencing temporary technical difficulties with the remaining 10%. Again, every partial report will be followed by a final report as soon as the analysis is complete, unless this requirement is waived.

### **What type of genetic counseling does Correlagen provide?**

Correlagen's professional staff is available to help ordering physicians select the best testing strategy or discuss the meaning of results. We are pleased to set up "curbside consults" where useful to physicians. We also provide materials designed to help the patient better understand the test results and to serve as preparation for a consultation with a genetic counselor.

### **How can the CardioGeneScan be used for research?**

Starting January 2010, Correlagen will offer clinician-researchers the opportunity to add genes to the CGS that are hypothesized to be associated with familial cardiac disease. By correlating variants detected in such candidate or "research" genes to clinical phenotype, their significance for cardiac disease can be clarified or proven over time. The cost for this service will be modest.

Testing of research genes will not be billed as a clinical service, and sequencing results will not be included in the clinical results report. Test results for research genes will be made available to researchers in an anonymized fashion only, and no variant interpretation will be provided. If specifically approved by an IRB, patients may be contacted for further information by Correlagen, but patient-identifying information will not be shared with researchers. If clinical utility is proven for a research gene, this gene may be added to the clinical portion of the CGS.

Patients will have the opportunity to opt-out of research testing.

### **Can gene panels be customized?**

At this time, we are only offering the six fixed panels described above. Multiple panels can be ordered either simultaneously or sequentially. Testing for single genes remains based on Sanger sequencing and is available for selected genes.

### **Additional Questions**

Please contact Correlagen Client Services at (781) 647-0604 or [testing@correlagen.com](mailto:testing@correlagen.com)