

Journal of Visualized Experiments

Amino acid-level signal-to-noise analysis of genetic variation: A method for determining the likelihood of variant pathogenicity --Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58907R1
Full Title:	Amino acid-level signal-to-noise analysis of genetic variation: A method for determining the likelihood of variant pathogenicity
Keywords:	genetics; Genetic Testing; mutation; Signal to Noise; Topology; Variant; Variant of Unknown Significance; Whole Exome Sequencing
Corresponding Author:	Andrew P Landstrom, MD, PhD Duke University School of Medicine Durham, NC UNITED STATES
Corresponding Author's Institution:	Duke University School of Medicine
Corresponding Author E-Mail:	andrew.landstrom@duke.edu
Order of Authors:	Edward G Jones
	Andrew P Landstrom, MD, PhD
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US\$2,400)
Please indicate the city, state/province, and country where this article will be filmed . Please do not use abbreviations.	Durham, NC, US



ANDREW P. LANDSTROM, MD, PhD

Assistant Professor
Department of Pediatrics, Division of Cardiology
Duke University School of Medicine

Duke University Medical Center
Box 2652
Durham, North Carolina 27701

Email: andrew.landstrom@duke.edu
Office: (919) 684-3028
Fax: (919) 385-9329

October 4, 2018
Manuscript Submission Cover Letter
Attn: Editorial Board, *Journal of Visualized Experiments*

Dear Dr. Ronald Myers,

It is my distinct pleasure to submit this revised manuscript entitled "Amino acid-level signal-to-noise analysis of genetic variation: A method for determining the likelihood of mutation pathogenicity" to JoVE as the first revision of an invited methods paper (JoVE58108). We believe that we have all the requested editorial and reviewer-requested modifications and have a much more user friendly, comprehensive, and impactful manuscript. We have also enclosed the requested permission we have received to publish a previously-published, now modified figure.

As the corresponding author for this manuscript, I attest that 1) this manuscript is not under consideration elsewhere, 2) none of the paper's contents with the exception of abstracts have been previously published, 3) all authors have read and approved of the manuscript, 4) all authors agree to be accountable for all aspects of the work, and 5) all conflicts of interest have been fully disclosed.

Sincerely,

A handwritten signature in black ink that reads "Andrew Landstrom". The signature is fluid and cursive.

Andrew Landstrom MD, PhD
Department of Pediatrics, Division of Cardiology
Duke University School of Medicine



TITLE:

Determining the Likelihood of Variant Pathogenicity Using Amino Acid-level Signal-to-noise Analysis of Genetic Variation

AUTHORS AND AFFILIATIONS:

Edward G. Jones, MD¹, Andrew P. Landstrom, MD, PhD²

¹Department of Pediatrics, Baylor College of Medicine, Houston, Texas, United States

²Department of Pediatrics, Division of Cardiology, Duke University School of Medicine, Durham, North Carolina, United States

Corresponding Author:

Andrew Landstrom, MD, PhD

Phone: (919) 684-3028

Fax: (919) 385-9329

Email Address of Co-Author:

Edward G. Jones, MD (edward.jones@bcm.edu)

KEYWORDS:

Genetic Analysis, genetic testing, mutation, topology, variant of unknown significance, whole exome sequencing

SUMMARY:

Amino acid signal-to-noise analysis determines the prevalence of genetic variation at a given amino acid position normalized to background genetic variation of a given population. This allows for identification of variant “hotspots” within a protein sequence (signal) that rises above the frequency of rare variants found in a population (noise).

ABSTRACT:

Advancements in the cost and speed of next generation genetic sequencing have generated an explosion of clinical whole exome and whole genome testing. While this has led to increased identification of likely pathogenic mutations associated with genetic syndromes, it has also dramatically increased the number of incidentally found genetic variants of unknown significance (VUS). Determining the clinical significance of these variants is a major challenge for both scientists and clinicians. An approach to assist in determining the likelihood of pathogenicity is signal-to-noise analysis at the protein sequence level. This protocol describes a method for amino-acid level signal-to-noise analysis that leverages variant frequency at each amino acid position of the protein with known protein topology to identify areas of the primary sequence with elevated likelihood of pathologic variation (relative to population “background” variation). This method can identify amino acid residue location “hotspots” of high pathologic signal, which can be used to refine the diagnostic weight of VUSs such as those identified by next-generation genetic testing.

INTRODUCTION:

The rapid improvement in genetic sequencing platforms has revolutionized the accessibility and role of genetics in medicine. Once confined to a single gene, or a handful of genes, the reduction in cost and increase in speed of next generation genetic sequencing has led routine sequencing of the entirety of the genome's coding sequence (whole exome sequencing, WES) and the entire genome (whole genome sequencing, WGS) in the clinical setting. WES and WGS has been used frequently in the setting of critically ill neonates and children with concern for genetic syndrome where it is a proven diagnostic tool that can change clinical management^{1,2}. While this has led to increased identification of likely pathogenic mutations associated with genetic syndromes, it has also dramatically increased the number of incidentally found genetic variants, or unexpected positive results, of unknown diagnostic significance. While some of these variants are disregarded and not reported, variants localizing to genes associated with potentially fatal or highly morbid diseases are often reported. Current guidelines recommend reporting of incidental variants found in specific genes which may be of medical benefit to the patient, including genes associated with the development of sudden cardiac death-predisposing diseases such as cardiomyopathies and channelopathies³. Although this recommendation was designed to capture individuals at risk of a SCD-predisposing disease, the sensitivity of variant detection far exceeds specificity. This is reflected in a growing number of VUSs and incidentally identified variants with unclear diagnostic utility that far exceed the frequency of the respective diseases in a given population⁴. One such disease, long QT syndrome (LQTS), is a canonical cardiac channelopathy caused by mutations localizing to genes which encode cardiac ion channels, or channel interacting proteins, resulting in delayed cardiac repolarization⁵. This delayed repolarization, seen by a prolonged QT interval on resting electrocardiogram, results in an electrical predisposition to potentially fatal ventricular arrhythmias such as torsades de pointes. While a number of genes have been linked to the development of this disease, mutations in KCNQ1-encoded I_{Ks} potassium channel (KCNQ1, Kv7.1) is the cause of LQTS type 1 and is utilized as an example below⁶. Illustrating the complexity in variant interpretation, the presence of rare variants in LQTS-associated genes, so called "background genetic variation" has been previously described^{7,8}.

In addition to large compendium-style databases of known pathogenic variants, several strategies exist for predicting the effect different variants will produce. Some are based on algorithms, such as SIFT and Polyphen 2, which can filter large numbers of novel non-synonymous variants to predict deleteriousness^{9,10}. Despite broad use of these tools, low specificity limits their applicability when it comes to "calling" clinical VUSs¹¹. "Signal-to-noise" analysis is a tool which identifies the likelihood of a variant being associated with disease based on the frequency of known pathologic variation at the loci in question normalized against rare genetic variation from a population. Variants localizing to genetic loci where there is a high prevalence of disease-associated mutations compared to population-based variation, a high signal-to-noise, are more likely to be disease-associated themselves. Further, rare variants found incidentally localizing to a gene with a high frequency of rare population variants compared to disease-associated frequency, a low signal-to-noise, may be less likely to be disease-associated. The diagnostic utility of signal-to-noise analysis has been illustrated in the latest guidelines for genetic testing for cardiomyopathies and channelopathies; however, it has only been employed at the whole gene level or domain-specific level¹². Recently, given increased availability of both pathologic variants

(disease databases, cohort studies in the literature) and population-based control variants (Exome Aggregation Consortium, ExAC and the Genome Aggregation Database, GnomAD¹³), this has been applied the individual amino acid positions within the primary sequence of a protein. Amino acid-level signal-to-noise analysis has proven useful in categorizing incidentally identified variants in genes associated with LQTS as likely “background” genetic variation rather than disease-associated. Among the three major genes associated with LQTS, including KCNQ1, these incidentally identified variants lacked any significant signal-to-noise ratios. Furthermore, when protein-specific domain topology was overlaid against areas of high signal-to-noise, pathologic mutation “hotspots” localized to key functional domains of the proteins¹⁴. This methodology holds promise in determining 1) the likelihood a variant is disease- or population-associated and 2) identifying novel critical functional domains of a protein associated with human disease.

PROTOCOL:

1. Identify the Gene and Specific Splice Isoform of Interest

NOTE: Here, we demonstrate the use of Ensembl¹⁵ to identify the consensus sequence for the gene of interest which is associated with the pathogenesis of the disease of interest (*i.e.* KCNQ1 mutations are associated with LQTS). Alternatives to Ensembl include RefSeq via the National Center for Biotechnology Information (NCBI)¹⁶ and the University of California, Santa Cruz (UCSC) Human Genome Browser¹⁷ (see **Table of Materials**).

1.1. In the Ensembl homepage, select the species (*i.e.* human) in the dropdown menu, and enter the gene of interest acronym in the field (*i.e.* KCNQ1). Click “Go”

1.2. Select the link corresponding to the gene of interest (*i.e.* “KCNQ1 (Human Gene)”

1.3. Select the link corresponding to the transcript of interest ID of interest from the “Transcript table” (*i.e.* TranscriptID ENST00000155840.10, NM_000218 [RNA transcript], NP_000209 [protein product of RNA transcript]).

NOTE: Review of the pertinent literature is needed to ensure the correct transcript consensus sequence is selected.

1.4. Note the transcript-specific NM and NP identification numbers for future reference found in the “RefSeq” column of the “Transcript Table”.

1.5. Select the link associated with the NP ID number to open a new webpage from the NCBI Protein database.

1.6. Scroll down to the “Origin” section to obtain the protein (primary) sequence for the gene transcript of interest.

1.7. Scroll up to the “Features” section to obtain a list of the protein features (functional domains, binding domains, post-translational modification sites).

NOTE: This information can also be obtained via the NCBI Protein database or from primary sources in the literature. This will be further discussed in step 5.

2. Create the Experimental Genetic Variant Database (the “Signal”)

NOTE: Here, we demonstrate how to create a database of disease-associated variants in the gene of interest with the frequency of the disease-associated variants among individuals with the disease of interest. This database can take many forms and represents the “signal” (phenotype-positive genetic variation) which will be normalized against the control variant database. This can include 1) disease-associated variants for comparison against VUSs or to identify novel functional domains of the protein and/or 2) VUSs to compare against disease-associated variants to determine likelihood of pathogenicity. Disease-associated variants in KCNQ1 will be presented for illustration; however, the method is the same for analysis of incidentally-identified VUSs or any other set of experimental variants.

2.1. Identify cohort(s) of unrelated index cases/probands with the disease of interest for which the gene of interest was comprehensively genotyped for all probands (*i.e.* a study identifies 24 unrelated probands hosting variants in KCNQ1 out of 200 individuals with LQTS who were subjected to KCNQ1 genetic interrogation).

NOTE: These cohorts can be identified from the literature, from experimental genetic analysis, or a combination of both.

2.1.1. Exclude studies which are not cohort-based (*i.e.* a case report describing a single mutation-positive individual), do not provide the total number of individuals genotyped for the gene of interest, or do not comprehensively genetically analyze the gene (*i.e.* a “targeted” genetic screening of only KCNQ1 exons 2-4).

2.1.2. Include individuals who are unrelated probands and exclude related individuals as this can over-estimate variant frequencies (*i.e.* a study identifies 4 unrelated individuals with KCNQ1 mutations in a cohort of 20 patients with LQTS. One of these probands is part of a family with 5 other mutation-positive kindred. Exclude all family members and include only the 4 unrelated probands).

2.2. Compile all experimental genetic variants found in identified cohort(s)

2.2.1. Assign nomenclature that contains the wild-type amino acid, amino acid position, and variant amino acid (*i.e.* alanine at amino acid number 212 changed to valine, A212V). One such type of nomenclature is demonstrated in **Figure 1**.

2.2.2. Confirm that variant nomenclature of all experimental genetic variants is based on the same reference gene transcript as noted in step 1.4. If experimental genetic variants are not annotated on the same reference gene transcript, then reannotate variant position to a reference

transcript using transcript alignment (see step 1.2)

2.3. Exclude variants that are not applicable depending on the question being explored.

2.3.1. Exclude variants localizing to non-coding regions of the genome or variants which do not alter the protein sequence such as synonymous, intronic variants, 5' or 3' untranslated region [UTR], and intergenic region variants (*i.e.* a reported pathologic variant in KCNQ1 which localizes to the 5' UTR of the coding region would be excluded as it is not predicted to alter the protein sequence).

2.3.2. Exclude variants that do not meet inclusion criteria for the study. For disease-associated variants, this includes variants that are no longer deemed pathologic.

2.3.2.1. Confirm that each variant is currently considered pathogenic, likely pathogenic, or at least not benign, by cross-referencing variants with the ClinVar database (see **Table of Materials**).

2.3.2.2. Enter the gene and variant of interest into ClinVar search field (*i.e.* KCNQ1-Y111C), select "Search"

2.3.2.3. Identify the variant of interest under the "Variation/Location" column.

2.3.2.4. Note the consensus interpretation of pathogenicity under "Clinical Significance" column (*i.e.* KCNQ1-Y111C is interpreted as "pathogenic").

2.3.2.5. Include variants which are "likely pathogenic" or "pathogenic."

2.3.2.6. Include variants with designations of "conflicting interpretations of pathogenicity," "uncertain significance," or when no record is available ("not provided") if warranted by the study.

2.3.2.7. Exclude variants designated as "likely benign" (*i.e.* KCNQ1-A62T).

2.4. Calculate the minor allele frequency (MAF) of each experimental variant position.

2.4.1. Calculate how many alleles were positive for each respective variant (*i.e.* if a KCNQ1-Y111C heterozygous mutation is found in 2 unrelated individuals, the number of variant-positive alleles is 2).

2.4.2. Calculate the total number of alleles sequenced within the cohort

2.4.2.1. Note the total number of individuals sequenced in each cohort study (step 2.1)

2.4.2.2. Multiply the total number of individuals by 2 to determine the total number of alleles

2.4.3. Calculate the total number of variant-positive individuals for each amino acid position (alleles in step 2.4.1/alleles in step 2.4.2). For example, if 2 unrelated individuals each host heterozygous KCNQ1-Y111C mutations in cohorts of 100 and 200 LQTS-afflicted individuals, respectively, then the frequency of experimental variants at amino acid position 111 is 2 variants/((100+200 individuals)*2 alleles/individual) (*i.e.* combined MAF 0.0033).

2.4.4. Calculate this value for each variant as the respective MAF of each experimental variant. For additional details see step 4.2.

3. Create the Control Genetic Variant Database (the “Noise”)

NOTE: Here, we demonstrate how to create a database of control variants in the gene of interest with an associated frequency in a control population. This database represents the “noise” (phenotype-negative, population-based genetic variation) which is the background against which the experimental variant database will be normalized. This is referred to as “control” variation.

3.1. Identify a cohort(s) of healthy, unrelated probands or utilize large population-based studies to identify rare variants among a given population.

NOTE: Sources for this database are diverse and include: 1) healthy individuals and/or otherwise phenotype-negative individuals subjected to Sanger sequencing, or publicly held databases of population-based individuals for which the disease in question is rare in frequency such as 2) 1000 Genome Project (N = 1,094 subjects)¹⁸, 3) National Heart, Lung, and Blood Institute GO Exome Sequencing Project (ESP, N = 5,379 subjects)¹⁹, 4) Exome Aggregation Consortium (ExAC, N = 60,706 subjects)¹³, and/or 5) Genome Aggregation Database (GnomAD, N = 138,632 individuals)¹³ (see **Table of Materials**). The GnomAD database will be utilized as an illustrative example.

3.1.1. Enter the gene of interest in the search box on the GnomAD homepage (*i.e.* KCNQ1).

3.1.2. Confirm that the browser selected the correct gene and transcript of interest (step 1.4).

3.1.3. Confirm that there is appropriate coverage of sequencing of the locus by reviewing “mean coverage” and “coverage plot.”

3.1.4. Select for coding sequence genetic variation by selecting “Missense + LoF.”

3.1.5. Select “Export table to CSV,” which will generate a TextEdit file named “Unknown.”

3.1.6. Relabel the file and include a new extension “*.csv” (*i.e.* “KCNQ1 Control Variation.csv”).

3.1.7. Open the file using an appropriate software program for analysis of *.csv files (see **Table of Materials**).

3.2. Identify the protein changing genetic variation in the column labeled "Protein Consequence."

3.3. Apply same exclusion criteria to these control genetic variants as the experimental genetic variants (step 2.3.1).

3.4. Identify the MAF of each control variant.

3.4.1. Locate the "Allele Count" column, which denotes the number of alleles found to harbor the variant.

3.4.2. Locate the "Allele Number" column, which denotes the total number of alleles sequenced at this given amino acid position.

NOTE: The total number of alleles sequenced will vary depending on coverage at that location. Areas of high coverage will approach 2 alleles*138,632 individuals (277,264 total alleles genotyped) while areas of lower coverage will have a reduced total allele number

3.4.3. Locate the variant MAF which is pre-calculated in the "Allele Frequency" column and represents "Allele Count" divided by "Allele Number."

NOTE: Human genomes have two of each allele (*i.e.* 1 subject found to have a heterozygous variant in 10 people has a MAF of 1/20)

3.4.4. Note the MAF for each variant as the respective MAF of each control variant.

NOTE: Variant specific MAF for each racial/ethnic group comprising GnomAD can be seen in the columns to the right of "Allele Frequency."

3.5. Apply a MAF threshold for rare variants above which control variants are excluded as "common."

3.5.1. Set the MAF threshold to the maximal value at which all truly disease-associated, pathologic variants (see step 2) that are also seen in the control database are included below the threshold (*i.e.*, among all disease-associated KCNQ1 variants also found in GnomAD the highest common variant MAF is 0.009, then all GnomAD variants above a threshold of 0.01 should be excluded).

3.6. Ensure that the experimental variant nomenclature is identical to control (see step 2.2).

3.7. Save the file. In some cases, this may require changing the file type/extension.

4. Amino Acid Level Signal-to-noise Calculation and Mapping

4.1. Calculate a MAF for each amino acid position with a control variant (see **Figure 1** containing example KCNQ1 GnomAD variants).

4.1.1. In a graphing-capable spreadsheet, create a column of the positions of all experimental variants.

4.1.2. Remove variant text to leave only the variant position.

NOTE: Various functions/formulas can be utilized to automatically (**Figure 1**, column C; see **Table of Materials**).

4.1.3. Sort the variants in ascending value to identify which positions have more than 1 variant associated with it (**Figure 1**, column E; *i.e.* amino acid position 10 is listed twice in column E which denotes 2 unique variants at the position).

4.1.4. Combine the MAF for each variant associated with a given position by taking the sum of all MAFs for a given position (**Figure 1**, column G and H).

4.2. Calculate a MAF for each amino acid position with an experimental variant (see **Figure 2** containing mock KCNQ1 pathologic variants).

4.2.1. In a similar fashion to 4.1.1, create a column of amino acid positions which have experimental variants (**Figure 2**, column B).

4.2.2. For each variant position, calculate the MAF of all variants associated with that position from step 2.4 (**Figure 2**, column C-G).

4.3. Create a rolling average of MAF for both experimental and control variants.

4.3.1. Expand the columns created in 4.1 and 4.2 to include cells for amino acid positions that have no variant as a MAF = 0. (**Figure 3**).

4.3.1.1. Create a column containing all amino acid positions in the gene of interest (*i.e.* 1 to 676 for KCNQ1, **Figure 3**, column C and I).

4.3.1.2. Add a MAF of 0 for all positions that do not have variants for both control and experimental data sets.

NOTE: This can be done automatically by utilizing the “VLOOKUP” function in a commonly utilized software program (**Figure 3**, column D and J, see **Table of Materials**).

4.3.2. Create a rolling average for each experimental and control prevalence column.

4.3.2.1. Create a column representing a rolling average of the MAF for both the for both control

and experimental data sets (**Figure 3**, column E and K).

4.3.2.2. In the rolling average column, place the average of the respective MAF for the 5 variant positions N-terminal and 5 variant positions C-terminal to the given position.

NOTE: This creates a rolling average of +/- 5. For positions with less than 5 amino acid residues preceding, or following, a rolling average location (*i.e.* the N- or C-terminus), the rolling average will only take into account those residues that are present (*i.e.* the rolling average at amino acid residue 3 will be an average of the MAF at amino acid residues 1 through 8, calculated as the sum of these MAFs divided by 8).

4.3.2.3. Calculate the cohort minimum frequency by dividing 1 by the total number of alleles in the control cohort (*i.e.* 1 divided by the number of individuals in the GnomAD control cohort (138,632) = 7.2×10^{-6}).

4.3.2.4. Change any cell with a control MAF of 0 to the minimum frequency to avoid dividing by 0 when calculating a signal-to-noise ratio.

4.4. Calculate the amino acid level signal-to-noise ratio (**Figure 4**).

4.4.1. Divide each amino acid position experimental rolling average by the respective control rolling average.

4.4.2. Graph this ratio (Y-axis) vs. amino acid position (X-axis).

5. Protein Domain Topology Overlay

5.1.1. Identify the consensus amino acid locations of functional domains/features, or areas of post-translational modification, of the protein of interest (step 1.7).

NOTE: A number of resources can be utilized to identify these domains. These resources, as well as resources for identifying putative domains in novel proteins, have been well reviewed in the literature²⁰. This protocol will describe the protein database available through NCBI, which is widely utilized and robust (see **Table of Materials**).

5.2. Identify amino acid positions associated with protein domains/features.

5.2.1. Open the NCBI webpage.

5.2.2. Enter the NP of the protein of interest into the search field.

5.2.3. Identify known protein domains and features are catalogues under "Features."

5.2.4. Identify and note the domain name/type and amino acid positions.

5.2.5. Select the link corresponding to the feature to visualize the region on the protein of interest primary sequence.

5.3. Create a column containing the boundaries of the domains/features.

5.3.1. Create a column next to the signal:noise column so that the amino acid position column can be referenced (Figure 5A, column C).

5.3.2. Identify the cells corresponding at the N-terminal or C-terminal aspect of each domain/feature and place a 1 in each cell (*i.e.* if the N-terminal domain of the S1 transmembrane domain of KCNQ1 is amino acid position 122, and the C-terminal domain is position 142, then a 1 is placed in the row for amino acid position 122 and 142).

5.3.3. For overlapping domains/features, display multiple domains by changing the 1 to other values (*i.e.* 1.5, 2, 2.5); this can assist in distinguishing domains.

5.4. Create a graph with these boundaries as a Y-axis and amino acid position on the X-axis (Figure 5B).

5.5. Overlay this graph with the signal-to-noise graph created in step 4.4.

5.6. Identify correlations between known protein domains/features and the signal-to-noise analysis.

6. Variant Position Overlay

6.1. Map individual variant positions for overlay of graphs produced in steps 4.4 and 5.4.

6.1.1. Create a column next to the domain/feature column such that rows in the column will correspond to amino acid positions (Figure 5A, column D).

6.1.2. Place a 1 in each cell in the added row corresponding to a position containing a respective variant.

6.1.3. Create a graph with this column as a Y-axis and amino acid position on the X-axis (Figure 5C).

6.2. Overlay this graph with the signal-to-noise graph created in step 4.4 and domain graph created in step 5.4.

REPRESENTATIVE RESULTS:

A representative result for amino acid-level signal to noise analysis for KCNQ1 is depicted in

Figure 6. In this example, rare variants identified in the GnomAD cohort (control cohort), incidentally-identified WES variants (experimental cohort #1), and LQTS case-associated variants deemed likely disease-associated (experimental cohort #2) are depicted. Further, the signal-to-noise analysis comparing the WES and LQTS cohort variant frequency normalized against GnomAD variant frequency is depicted. LQTS-associated variants demonstrated high signal-to-noise ratios in domains corresponding with the channel pore, selectivity filter, and the KCNE1-binding domain. In comparison, incidentally identified variants in the WES cohort did not clearly demonstrate specific regions of high signal-to-noise elevation, suggesting that these variants reflect background genetic variation. This example did not utilize variant MAFs as noted above; however, it demonstrates all of the same principles as described.

FIGURE AND TABLE LEGENDS:

Figure 1: Example of control variant database with MAF calculation. Column A, directly imported GnomAD control rare variants. Column B, deletion of right-sided, non-position-related text from the variant nomenclature using an example formula for character removal (*i.e.*: for B2 “=RIGHT(A2,LEN(A2)-5”, see **Table of Materials**). Column C, deletion of left-sided, non-position-related text from the variant nomenclature using a related formula (*i.e.*: for C2 “=LEFT(B2,LEN(B2)-3”). Column D, resultant unsorted amino acid positions. Column E, amino acid positions sorted in an ascending fashion to allow for identification of duplicate positions. Column F, associated MAF for each variant as imported from GnomAD. Column G and H, combined MAF for a given amino acid position (sum of each variant MAF at a specific position).

Figure 2: Example of experimental variant database with MAF calculation. Column A, a list of mock LQTS-associated mutations in KCNQ1 representing a disease-associated mutation experimental database. Column B, mutation position corresponding to each variant. Column C, a count of mutation-positive individuals within mock Study 1. Each are presumed to be heterozygous mutation carriers. The total number of individuals genotyped in the study is located at the bottom of the sheet. Column D, count of mutation-positive individual in mock Study 2. Column E, count of mutation-positive individual in mock Study 3. Column F, total mutation-positive individuals hosting the observed mutation across all studies. Note that distinct mutations associated with the same amino acid position should be combined. Column G, MAF of each mutation and amino acid position using an example formula (*i.e.*: for G2 “=2/(176*2)”, see **Table of Materials**). Note that since all individuals are presumed to be heterozygous and each individual presumed to carry 2 alleles of the KCNQ1 locus, the total individuals should be multiplied by 2 for the allele frequency.

Figure 3: Example of rolling average calculation for control and experimental variants. Column A and B, GnomAD control variant positions and respective MAFs. Column C, all amino acid positions of KCNQ1 from amino acid position to final. Column D, GnomAD variant MAF for all positions with a MAF of 0 in place of positions without a variant. This can be automatically calculated using a VLOOKUP function (*i.e.* for D2, “=IFERROR(VLOOKUP(C2,A:B,2,0),0”, see **Table of Materials**). Column E, rolling average of position MAF using an example formula (*i.e.* for E2, “=SUM(D2:D7)/6” and for E7, “=SUM(D2:D12)/11”). Column G and H, LQTS experimental variant

positions with respective MAFs. Column I, all amino acid positions of KCNQ1. Column J, LQTS variant MAF for all positions. Column K, rolling LQTS MAF. Gray fill cells are examples of where MAF values from columns B and H are expanded into column D and J, respectively, which correlate with respective positions in column C/I. Note that it is critical that all cells are formatted as “Numbers” for proper formula functioning.

Figure 4: Example of signal-to-noise analysis and graphing. Left, example database and calculations. Column A, all amino acid positions of KCNQ1. Column B, LQTS experimental MAF rolling average for each position. Column C, GnomAD control MAF rolling average for each position. D: Signal-to-noise ratio (*i.e.* for D2, “=B2/C2”). Right, example of graph of signal-to-noise ratio (Y-axis) versus amino acid position (X-axis).

Figure 5: Example of protein and variant position mapping. A, example database and calculations. Column A, all amino acid positions of KCNQ1. Column B, KCNQ1 positions which have a rare control variant identified in GnomAD. Column C, the domain mapping column where cells containing values correspond to the N or C-terminal aspect of identified KCNQ1 protein domains or features. As the most N-terminal domain is the S1 domain has the N-terminal boundary at amino acid 122, no values are noted here. Column D, the variant mapping column where cells containing a 1 correspond to KCNQ1 positions which localize rare variants. Gray fill cells are two examples of where variant positions in column B are expanded into column D which correlate with respective positions in column A.

Figure 6: Example of amino acid-level signal-to-noise analysis of KCNQ1-encoded KCNQ1 (Kv7.1). Top, variant positions are demonstrated with vertical lines, including rare GnomAD cohort variants (black), incidentally-identified variants in WES referrals (blue), and variants identified in LQTS cases (green). Functional domains are noted. Relative frequency of LQTS case variants normalized to GnomAD variants (green line) is depicted compared to WES (blue line). S1-S6, transmembrane domains; SF, ion selectivity filter; KCNE1 and AKAP9, respective protein binding domains. Modified and reprinted with permission from previous work¹⁴.

DISCUSSION:

High-throughput genetic testing has advanced dramatically in its application and availability over the past decade. However, in many diseases with well-established genetic underpinnings, such as cardiomyopathies, expanded testing has failed to improve diagnostic yield²¹. Further, there is significant uncertainty regarding the diagnostic utility of many identified variants. This is partially due to a growing number of incidentally identified rare variants discovered on WES and WGS, which can lead to misdiagnosis²². Amino acid level signal-to-noise analysis is based on well-established strategies for predicting variant pathogenicity and provides the advantage of leveraging large-scale population-based genome studies to refine variant interpretation.

It follows that one of the most crucial steps to this protocol is the selection of control and experimental cohorts. Many of the publicly-available large genome studies are accessible through aggregate databases, such as GnomAD, that can allow for representative control cohorts

in this protocol to be as large as 138,632 individuals at present date. Though not all subjects in these aggregate cohorts are ostensibly healthy, the large sample size in the setting of rare disease makes this resource invaluable and allows for a stringent MAF exclusion threshold. Exclusion of common variants is necessary as they are unlikely to be a cause of highly penetrant Mendelian disease. Based on previous work, a MAF threshold of 0.01 for channelopathy-associated genes and 0.0001 for cardiomyopathy genes may be appropriate and has been validated by independent groups^{23,24}. Importantly, given the importance of the MAF threshold, this should be set and validated for each study independently. A MAF threshold need not be applied to an experimental cohort, given the well-established presence of founder mutations in channelopathies and cardiomyopathies. The size of the experimental cohort needs to be sufficient to identify areas where variants may cluster; however, there is no strict size. Additionally, the experimental cohort should not include variants known to be benign within the literature, as this would diminish the veracity of the pathogenic signal.

Properly selecting exclusion criteria is also crucial for interpretation and applicability of the result. Though this protocol recommends excluding certain mutation classes such as synonymous variants, these could feasibly be included for disease processes in which deleterious synonymous variants have been identified^{25,26}. Additionally, when various exclusion criteria are applied to both experimental and control groups, it can allow for stratification of signal-to-noise mapping by mutation subclass (*i.e.* comparing missense to truncating variants).

Setting a rolling average for MAFs allow for inference of involvement to neighboring amino acids. For example, if amino acid position 35 contains a pathologic variant and resides in a critical protein domain, then position 36 may have a degree of pathogenicity when mutated. Likewise, should a stretch of primary sequence have a large amount of rare control variants, then amino acids within this region that do not host rare variants may yet have a higher likelihood of containing rare variants found in a population. While the rolling average in this protocol is +/- 5, this range can be vary based on the user's desired level of resolution of signal-to-noise ratio and the specific protein being studied. In the example of LQTS, the interrogated KCNQ1-encoded KCNQ1 channel has several transmembrane domains spanning ~10 amino acids, prompting the authors to adjust their desired resolution to reflect significant findings on that scale¹⁴. For proteins with a longer primary sequence and protein length, the span of the rolling average may need to be increased due to larger spans of protein sequence without control variation.

There are several limitations to this method. As previously stated, a sufficient phenotype-positive population hosting putative pathologic variants must be identified in order to drive a clear pathologic signal. Additionally, these pathologic variants may have variable penetrance, thus truly pathologic mutations may not manifest a disease phenotype or may otherwise not be fully penetrant and disease causing. While many publicly held databases, such as GnomAD, are often considered "healthy cohorts," the prevalence of genetic diseases is likely similar in this database as population studies. As detailed, this protocol focuses specifically on amino-acid level changes resulting from exonic gene variants that code for amino acids, which excludes the role that pathogenic intronic splicing variants may play in monogenic disease. Given their recently demonstrated role in cardiomyopathies, expansion of the resolution this approach may be

warranted to identify intergenic “hotspots” as well. Furthermore, the application of a MAF threshold may miss certain “risk alleles” that, though existing in the population with a MAF higher than that of disease prevalence, may contribute to disease pathogenesis^{27,28}. Despite these limitations, this analysis is adaptable and can play a key role in providing clinicians a relative probability of disease pathogenicity when appropriate applied.

Finally, given the predilection of this analysis to identify critical regions within a protein, amino acid-level signal-to-noise calculations utilizing pathologic mutations offers the possibility of identifying novel functional domains of the proteins being studied. Given the observation of high pathogenicity signal-to-noise at key locations of ion channels, such as the pore domain, selectivity filter, S2 transmembrane domain, and the KCNE1-binding domain of KCNQ1, identification of a “peak of pathogenicity” within an area of the protein without a known function may suggest a novel critical domain. For example, a marked peak of pathogenicity of LQTS-associated mutations has been identified localizing to amino acid residues 912-930 of KCNH2-encoded KCNH2 (Kv11.1). This region of the protein has no identifiable functional domain yet demonstrates a marked propensity for LQTS-associated mutations¹⁴. As the knowledge of protein topology expands, more sophisticated proteomics could feasibly improve the resolution of this method in the future from analyzing signal-to-noise ratio along a protein’s primary structure to include its secondary, tertiary, or quaternary structure. Addition of advanced computational sciences to this analysis, such as machine learning and artificial intelligence, affords the opportunity to identify novel patterns among pathologic versus population-based genetic variation, if robust databases of these variants can be generated^{29,30}. In turn, this method could aid in better characterizing and predicting the genotype-phenotype relationship of specific diseases and be used in conjunction with an individual’s pre-test probability of disease to improve the diagnostic yield of genetic testing. Further, this analysis may discover novel protein biology and identify novel loci within the human genome which manifest with disease when altered.

ACKNOWLEDGMENTS:

APL is supported by the National Institutes of Health K08-HL136839.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

- 1 Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *New England Journal of Medicine*. **369** (16), 1502-1511, (2013).
- 2 Meng, L. *et al.* Use of Exome Sequencing for Infants in Intensive Care Units: Ascertainment of Severe Single-Gene Disorders and Effect on Medical Management. *Journal of the American Medical Association Pediatrics*. **171** (12), e173438, (2017).
- 3 Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine*. **19** (2), 249-255, (2017).

- 4 Landstrom, A. P. & Ackerman, M. J. The Achilles' heel of cardiovascular genetic testing: distinguishing pathogenic mutations from background genetic noise. *Clinical Pharmacology and Therapeutics*. **90** (4), 496-499, (2011).
- 5 Landstrom, A. P., Tester, D. J. & Ackerman, M. J. Role of genetic testing for sudden death predisposing heart conditions in athletes. In Lawless C. (eds) *Sports Cardiology Essentials*. Springer, New York, NY (2011).
- 6 Wang, Q. *et al.* Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nature Genetics*. **12** (1), 17-23, (1996).
- 7 Kapa, S. *et al.* Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. *Circulation*. **120** (18), 1752-1760, (2009).
- 8 Ackerman, M. J. *et al.* Ethnic differences in cardiac potassium channel variants: implications for genetic susceptibility to sudden cardiac death and genetic testing for congenital long QT syndrome. *Mayo Clinic Proceedings*. **78** (12), 1479-1487, (2003).
- 9 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. **4** (7), 1073-1081, (2009).
- 10 Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*. **Chapter 7** (Unit 7.20), (2013).
- 11 Flanagan, S. E., Patch, A. M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic Testing and Molecular Biomarkers*. **14** (4), 533-537, (2010).
- 12 Ackerman, M. J. *et al.* HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart Rhythm*. **8** (8), 1308-1339, (2011).
- 13 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. **536** (7616), 285-291, (2016).
- 14 Landstrom, A. P. *et al.* Amino acid-level signal-to-noise analysis of incidentally identified variants in genes associated with long QT syndrome during pediatric whole exome sequencing reflects background genetic noise. *Heart Rhythm*. **15** (7), 1042-1050, (2018).
- 15 Hubbard, T. *et al.* Ensembl 2005. *Nucleic Acids Research*. **33** (Database issue), D447-453, (2005).
- 16 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*. **44** (D1), D733-745, (2016).
- 17 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Research*. **12** (6), 996-1006, (2002).
- 18 The 100 Genome Projects Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. **491** (7422), 56-65, (2012).
- 19 Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*. **493** (7331), 216-220, (2013).
- 20 Mulder, N. J. & Apweiler, R. Tools and resources for identifying protein families, domains and motifs. *Genome Biology*. **3** (1), reviews2001.1-reviews2001.8, (2002).

- 21 Cirino, A. L. *et al.* A Comparison of Whole Genome Sequencing to Multigene Panel Testing in Hypertrophic Cardiomyopathy Patients. *Circulation Cardiovascular Genetics*. **10** (5), (2017).
- 22 Landstrom, A. P. *et al.* Interpreting Incidentally Identified Variants in Genes Associated With Catecholaminergic Polymorphic Ventricular Tachycardia in a Large Cohort of Clinical Whole-Exome Genetic Test Referrals. *Circulation Arrhythmia and Electrophysiology*. **10** (4), (2017).
- 23 Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genetics in Medicine*. **19** (10), 1151-1158, (2017).
- 24 Walsh, R. *et al.* Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genetics in Medicine*. **19** (2), 192-203, (2017).
- 25 Buske, O. J., Manickaraj, A., Mital, S., Ray, P. N. & Brudno, M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics*. **31** (5), 799, (2015).
- 26 Wen, P., Xiao, P. & Xia, J. dbDSM: a manually curated database for deleterious synonymous mutations. *Bioinformatics*. **32** (12), 1914-1916, (2016).
- 27 Bagnall, R. D. *et al.* Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. *Journal of the American College of Cardiology*. **72** (4), 419-429, (2018).
- 28 Giudicessi, J. R., Roden, D. M., Wilde, A. A. M. & Ackerman, M. J. Classification and Reporting of Potentially Proarrhythmic Common Genetic Variation in Long QT Syndrome Genetic Testing. *Circulation*. **137** (6), 619-630, (2018).
- 29 Sundaram, L. *et al.* Predicting the clinical impact of human mutation with deep neural networks. *Nature Genetics*. **50**, 1161-1170, (2018).
- 30 Krittanawong, C., Zhang, H., Wang, Z., Aydar, M. & Kitai, T. Artificial Intelligence in Precision Cardiovascular Medicine. *Journal of the American College of Cardiology*. **69** (21), 2657-2664, (2017).

Control Variants

	A	B	C	D	E	F	G	H
1	Protein Consequence	Position	Position	Unsort	Sort	MAF	Position	Combined MAF
2	p.Pro75Ser	75Ser	7	7	3	3.24E-05	3	3.24E-05
3	p.Ala55Val	55Val	55	55	5	3.94E-05	5	3.94E-05
4	p.Ala62Val	62Val	62	62	6	5.64E-05	6	5.64E-05
5	p.Ser66Tyr	66Tyr	66	66	7	0.0001495	7	0.0001495
6	p.Pro67Leu	67Leu	67	67	10	3.25E-05	10	6.23E-05
7	p.Pro73Thr	73Thr	73	73	10	2.98E-05		
8	p.Pro73Ser	73Ser	73	73	16	1.54E-05	16	1.54E-05
9	p.Ala76Thr	76Thr	76	76	17	1.45E-05	17	1.45E-05
10	p.Ser77Phe	77Phe	77	77	19	1.45E-05	19	1.45E-05
11	p.Asp88Glu	88Glu	88	88	28	1.90E-05	28	2.11E-04
12	p.Pro89Thr	89Thr	89	89	28	0.0001919		
13	p.Pro89Leu	89Leu	89	89	29	9.18E-05	29	9.18E-05
14	p.Val91Leu	91Leu	91	91	30	3.23E-05	30	3.93E-05
15	p.Ser92Pro	92Pro	92	92	30	6.98E-06		
16	p.Ser92Cys	92Cys	92	92	32	6.98E-06	32	6.98E-06
17	p.Ile93Val	93Val	93	93	34	1.39E-05	34	1.39E-05
18	p.Tyr94ThrfsTer143	94ThrfsTer143	94ThrfsTer	94	35	3.49E-05	35	3.49E-05
19	p.Tyr94Cys	94Cys	94	94	37	3.44E-05	37	3.44E-05
20	p.Thr96Ala	96Ala	96	96	38	6.96E-06	38	6.96E-06
21	p.Thr96Arg	96Arg	96	96	39	5.57E-05	39	5.57E-05
22	p.Arg98His	98His	98	98	40	1.39E-05	40	1.39E-05
23	p.Pro99Arg	99Arg	99	99	45	0.0007161	45	7.30E-04
24	p.Thr104Ile	104Ile	104	104	45	1.39E-05		
25	p.Gln107Ter	107Ter	107	107	46	3.23E-05	46	3.23E-05
26	p.Val110Ile	110Ile	110	110	47	0.000126	47	0.000126
27	p.Tyr111Cys	111Cys	111	111	48	4.88E-05	48	9.05E-05
28	p.Cys122LeufsTer163	122LeufsTer163	122LeufsTer	122	48	4.18E-05		
29	p.His3Gln	3Gln	3	3	49	2.87E-05	49	2.87E-05
30	p.Ala5Gly	5Gly	5	5	50	3.48E-05	50	3.48E-05

Experimental Variants

	A	B	C	D	E	F	G
1	LQTS Mutation	Position	Study 1	Study 2	Study 3	Total	Mutation Freq
2	R12Q	12	1	1		2	0.022727273
3	W15L	15		1		1	0.011363636
4	G22F	22	3	4	2	9	0.102272727
5	A49S	49	1			1	0.011363636
6	G57T	57	1			1	0.011363636
7	G60E	60		1		1	0.011363636
8	A69P	69			1	1	0.011363636
9	R83C	83	1			1	0.011363636
10	R89M	89		1		1	0.011363636
11	V91W	91	1			1	0.011363636
12	V124T	124		2	1	3	0.034090909
13							
14							
15	Total Individuals		49	75	52	176	

Figure 3

Rolling Average

	A	B	C	D	E	F	G	H	I	J	K
	GnomAD		All		Gnom Roll		LQTS		All		LQTS Roll
1	Positions	GnomAD MAF	Positions	GnomAD MAF	MAF		Position	LQTS MAF	Positions	LQTS MAF	MAF
2	3	3.24E-05	1	0	2.14E-05		12	0.0227273	1	0	0.00E+00
3	5	3.94E-05	2	0	3.97E-05		15	0.0113636	2	0	0.00E+00
4	6	5.64E-05	3	3.24E-05	3.47E-05		22	0.1022727	3	0	0.00E+00
5	7	0.0001495	4	0	3.09E-05		49	0.0113636	4	0	0.00E+00
6	10	3.25E-05	5	3.94E-05	3.40E-05		57	0.0113636	5	0	0.00E+00
7	10	2.98E-05	6	5.64E-05	3.09E-05		60	0.0113636	6	0	0.00E+00
8	16	1.54E-05	7	0.0001495	3.09E-05		69	0.0113636	7	0	2.07E-03
9	17	1.45E-05	8	0	3.09E-05		83	0.0113636	8	0	2.07E-03
10	19	1.45E-05	9	0	2.80E-05		89	0.0113636	9	0	2.07E-03
11	28	1.90E-05	10	6.23E-05	2.80E-05		91	0.0113636	10	0	3.10E-03
12	28	0.0001919	11	0	2.58E-05		124	0.0340909	11	0	3.10E-03
13	29	9.18E-05	12	0	2.20E-05				12	0.0227273	3.10E-03
14	30	3.23E-05	13	0	8.37E-06				13	0	3.10E-03
15	30	6.98E-06	14	0	9.69E-06				14	0	3.10E-03
16	32	6.98E-06	15	0	9.69E-06				15	0.0113636	3.10E-03
17	34	1.39E-05	16	1.54E-05	4.03E-06				16	0	3.10E-03
18	35	3.49E-05	17	1.45E-05	4.03E-06				17	0	1.24E-02
19	37	3.44E-05	18	0	4.03E-06				18	0	1.03E-02
20	38	6.96E-06	19	1.45E-05	4.03E-06				19	0	1.03E-02

Figure 4

Signal to Noise Analysis

	A	B	C	D
	All	LQTS Roll	Gnom Roll	LQTS:GnomAD
1	Positions	MAF	MAF	(S:N)
2	1	0.00E+00	5.41E-06	0.00
3	2	0.00E+00	1.03E-05	0.00
4	3	0.00E+00	1.60E-05	0.00
5	4	0.00E+00	3.09E-05	0.00
6	5	0.00E+00	2.78E-05	0.00
7	6	0.00E+00	2.52E-05	0.00
8	7	2.07E-03	2.82E-05	73.27
9	8	2.07E-03	2.82E-05	73.27
10	9	2.07E-03	2.82E-05	73.27
11	10	3.10E-03	2.82E-05	109.91
12	11	3.10E-03	2.52E-05	122.74
13	12	3.10E-03	2.52E-05	122.74
14	13	3.10E-03	2.31E-05	134.35
15	14	3.10E-03	1.93E-05	160.97
16	15	3.10E-03	5.66E-06	547.29
17	16	3.10E-03	6.98E-06	444.18
18	17	1.24E-02	6.98E-06	1776.72
19	18	1.03E-02	4.03E-06	2566.31
20	19	1.03E-02	4.03E-06	2566.31
21	20	1.03E-02	4.03E-06	2566.31
22	21	9.30E-03	4.03E-06	2309.68
23	22	9.30E-03	4.03E-06	2309.68
24	23	9.30E-03	4.03E-06	2309.68
25	24	9.30E-03	2.63E-06	3536.40
26	25	9.30E-03	3.05E-06	3052.92
27	26	9.30E-03	1.14E-05	816.35
28	27	9.30E-03	1.30E-05	714.59
29	28	0.00E+00	1.30E-05	0.00
30	29	0.00E+00	1.36E-05	0.00
31	30	0.00E+00	1.36E-05	0.00

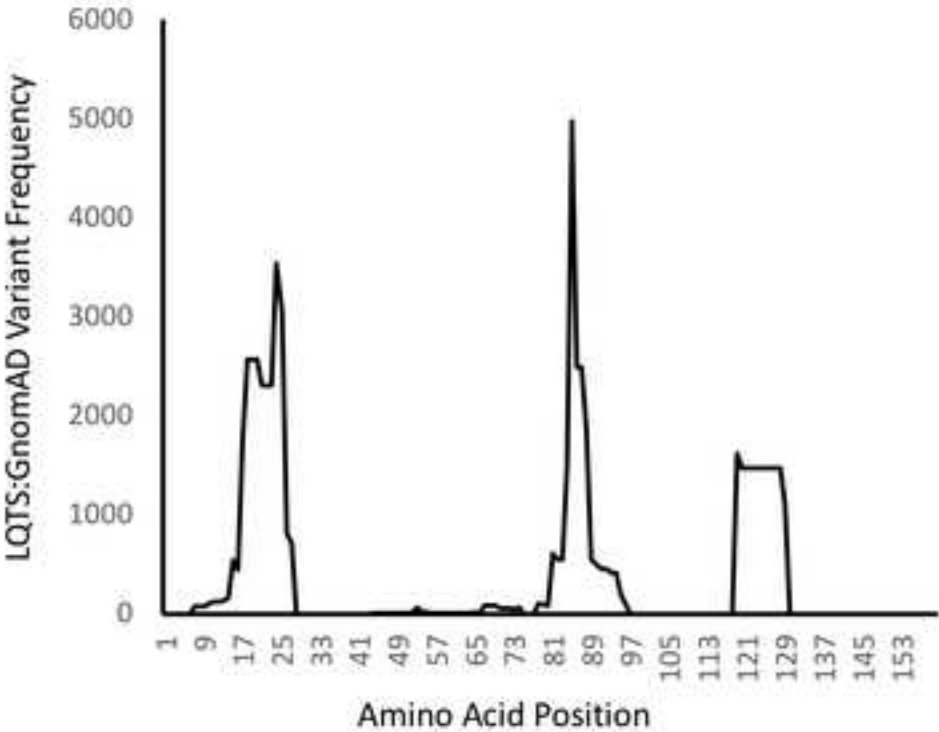
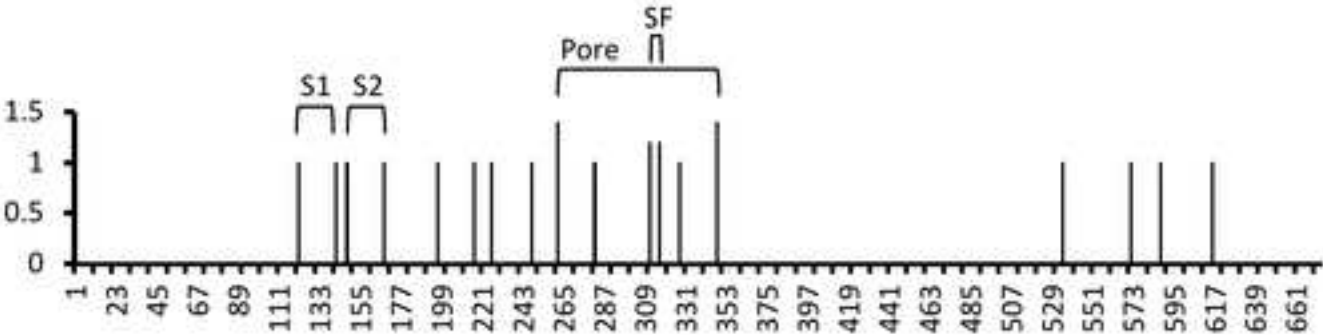


Figure 5

A Mapping

	A	B	C	D
	All KCNQ1	GnomAD		
1	Positions	Variant	Domain	1 = GnomAD
2		Position	Boundary	Variant
3	1	3		
4	2	5		
5	3	6		1
6	4	7		
7	5	10		1
8	6	10		1
9	7	16		1
10	8	17		
11	9	19		
12	10	28		1
13	11	28		
14	12	29		
15	13	30		
16	14	30		
17	15	32		
18	16	34		1
19	17	35		1
20	18	37		
21	19	38		1
22	20	39		

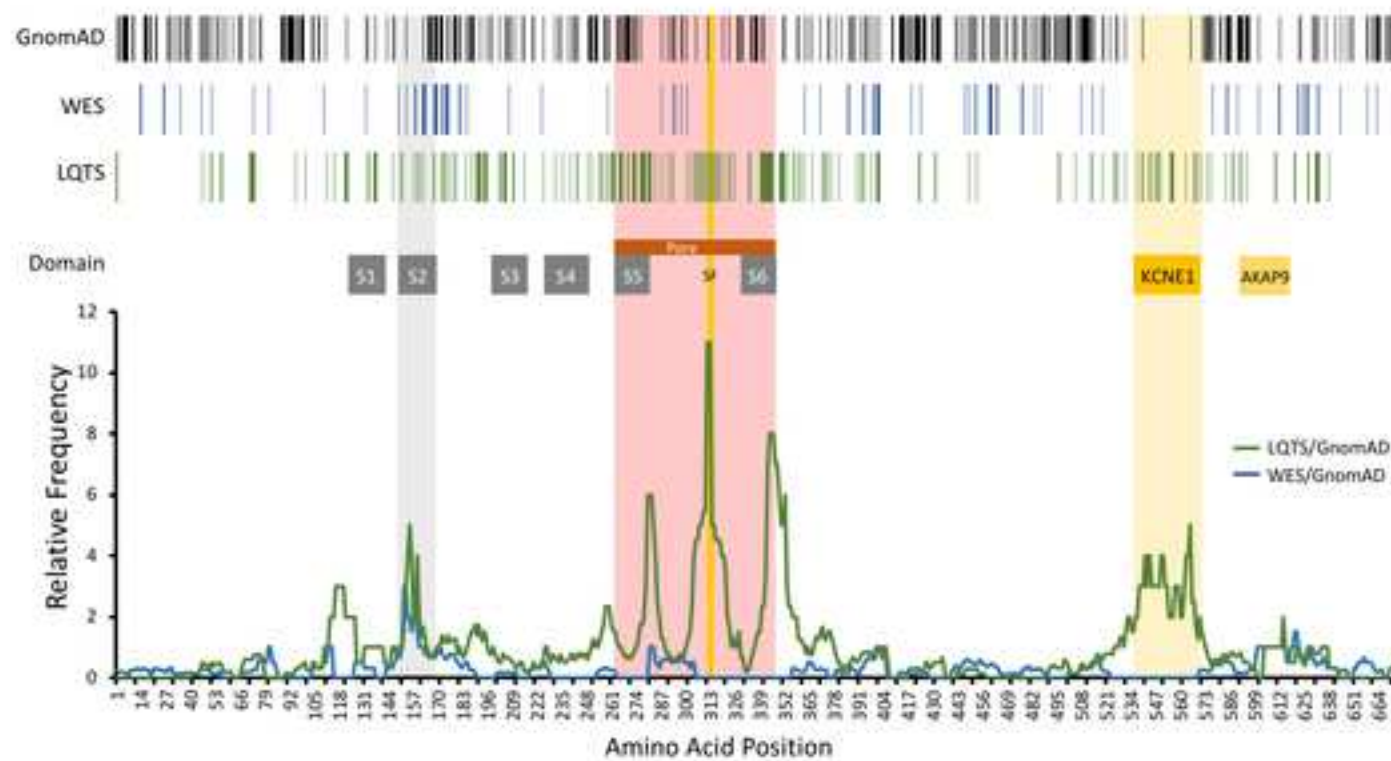
B Domain/Feature Mapping



C Variant Mapping



Figure 6



Name of Material/ Equipment	Company
1000 Genome Project	N/A
ClinVar	N/A
Ensembl Genome Browser	N/A
Excel	Microsoft
Exome Aggregation Consortium	N/A
Genome Aggregation Database	N/A
National Center for Biotechnology Information Domain and Structure Database	N/A
National Center for Biotechnology Information Gene Database	N/A
National Center for Biotechnology Information Protein Database	N/A
National Heart, Lung, and Blood Institute GO Exome Sequencing Project	N/A
SnapGene	GSL Biotech LCC
University of California, Santa Cruz Human Genome Browser	N/A

Catalog Number

www.internationalgenome.org

www.ncbi.nlm.nih.gov/clinvar

uswest.ensembl.org/index.html

office.microsoft.com/excel/

www.exac.broadinstitute.org

www.gnomad.broadinstitute.org

www.ncbi.nlm.nih.gov/guide/domains-structures/

www.ncbi.nlm.nih.gov/gene/

www.ncbi.nlm.nih.gov/protein/

www.evs.gs.washington.edu/EVS/

www.snapgene.com

www.genome.ucsc.edu

Comments/Description

Used for all example formulas and functions



1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Amino acid-level signal-to-noise analysis of genetic variation: A method for determining the likelihood of mutation pathogenicity

Author(s): Edward G. Jones and Andrew P. Landstrom

Item 1: The Author elects to have the Materials be made available (as described at <http://www.jove.com/publish>) via:

☒ Standard Access

☐ Open Access

Item 2: Please select one of the following items:

☒ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "Agreement" means this Article and Video License Agreement; "Article" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "Author" means the author who is a signatory to this Agreement; "Collective Work" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "CRC License" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>; "Derivative Work" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "Institution" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "JOVE" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "Materials" means the Article and / or the Video; "Parties" means the Author and JOVE; "Video" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JOVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion

of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JOVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JOVE agreeing to publish the Article, the Author hereby grants to JOVE, subject to Sections 4 and 7 below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in Item 1 above, JOVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

ARTICLE AND VIDEO LICENSE AGREEMENT

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum

rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to


the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US\$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to be one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Andrew P. Landstrom	
Department:	Pediatrics	
Institution:	Duke University School of Medicine	
Title:	Assistant Professor	
Signature:		Date: 8/1/18

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Response to Reviewers
JoVE Submission JoVE58907

We outline our point-by-point response to the editor 2 reviewers comments below. To aid in review of these revisions, we have included revised manuscript text in this document. Our comments are in bold.

Editorial Comments:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

Done. We have made several small changes to aid in readability.

2. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."

Figure 6 is a modified version of a figure initially published by our group in:

"Amino acid-level signal-to-noise analysis of incidentally identified variants in genes associated with long QT syndrome during pediatric whole exome sequencing reflects background genetic noise." Heart Rhythm 2018 Mar 2. pii: S1547-5271(18)30142-5. doi: 10.1016/j.hrthm.2018.02.031

Elsevier journal authors retain the right to use or re-use excerpts (including figures) in other works with proper acknowledgements. We have provided this acknowledgement in the figure legend. For the convenience of the editor, we have also attached the written communication with the Elsevier editorial desk and the appropriate policy.

3. Please provide an email address for each author.

Added

4. Please refrain from using indentations.

Removed

5. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents. You may use the generic term followed by "(see table of materials)" to draw the readers' attention to specific commercial names. Examples of commercial sounding language in your manuscript are: EXCEL, GSL Biotech LCC, etc.

Removed and the Table of Materials referenced and updated.

6. Please revise the protocol text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).

Removed.

7. Please revise the protocol to contain only action items that direct the reader to do something (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “Note.” Please include all safety procedures and use of hoods, etc. However, notes should be used sparingly and actions should be described in the imperative tense wherever possible.

We have extensively revised the protocol to achieve this. While we have significantly reduced the number of “Notes” that are included in the protocol portion of the manuscript, the ones that remain we feel are critical for usability of the protocol.

8. 2.1, 2.2, 2.3.4, 2.4.1, 2.4.2, 2.6, 3.1, 3.2, etc.: Please write the text in the imperative tense.

As above.

9. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.

As above. We have moved topics on MAF threshold and rolling average to the Discussion.

10. 5.1.3 and sub-steps: Please describe the specific actions being performed here. If there are no actions, they should be removed from the protocol.

Revised.

11. Please note that your protocol will be used to generate the script for the video and must contain everything that you would like shown in the video. Software must have a GUI (graphical user interface) and software steps must be more explicitly explained ('click', 'select', etc.). Please add more specific details (e.g. button clicks for software actions, numerical values for settings, etc.) to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol.

Done. We have added more functional details. Due to the revision that requires us to remove “Excel” from the manuscript (referencing the Table of Materials) this limits our ability to describe how to use this software if we must keep the software a generic description.

12. Please include single-line spaces between all paragraphs, headings, steps, etc.

Done

13. After you have made all the recommended changes to your protocol (listed above), please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the

essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol.

Done

14. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in imperative tense.

Done

15. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.

Done

16. References: Please do not abbreviate journal titles.

Done

Reviewer #1:

- 1) Point 4.3 Rolling average - further guidance could be provided as to what this should be - e.g. should it be 5? and what might this be based upon.

Thank you for pointing out our lack of clarity. It's true that we do not specify the range of this rolling average beyond the given example, nor do we provide guidance on factors that may explain (1) why 5 was chosen and (2) when a bigger or smaller range may be appropriate. +/- 5 was utilized in KCNQ1 given the size of the transmembrane domains (~10 amino acids) and needing to illustrate a signal to noise "resolution" that would be less than these domains. The rolling average can similarly be adjusted to suit 1) protein domain size, 2) length of the primary sequence, 3) the prevalence of rare variants found in healthy individuals, 4) the signal-to-noise graph itself. We address this with the following addition to the Discussion and have included relevant addition below.

"Setting a rolling average for MAFs allow for inference of involvement to neighboring amino acids. For example, if amino acid position 35 contains a pathologic variant and resides in a critical protein domain, then position 36 may have a degree of pathogenicity when mutated. Likewise, should a stretch of primary sequence have a large amount of rare control variants, then amino acids within this region that do not host rare variants may yet have a higher likelihood of containing rare variants found in a population. While the rolling average in this protocol is +/- 5, this range can be vary based on the user's desired level of resolution of signal-to-noise ratio and the specific protein being studied. In the example of LQTS, the interrogated *KCNQ1*-encoded KCNQ1 channel has several transmembrane domains spanning ~10 amino acids, prompting the authors to adjust their desired resolution to reflect significant findings on that scale¹. For proteins with a longer primary sequence and protein length, the span of the rolling average may need to be increased due to larger spans of protein sequence without control variation."

Reviewer #2:

- 1) Introduction - as this protocol uses long QT syndrome as the primary example, the authors ought to cite some of the sentinel work pertaining to background genetic noise this protocol expands upon (Ackerman, Mayo Clinic Proceedings, 2003, Kapa et al, Circulation, 2009, etc).

Thank you for pointing out this oversight, as these works contributed heavily to our current work. Citations for these papers have been added. The introduction now states:

While a number of genes have been linked to the development of this disease, mutations in *KCNQ1*-encoded I_{Ks} potassium channel (*KCNQ1*, Kv7.1) is the cause of LQTS type 1 and is utilized as an example below². Illustrating the complexity in variant interpretation, the presence of rare variants in LQTS-associated genes, so called “background genetic variation” has been previously described^{3,4}.

- 2) In light of the recent work by Bagnall et al, JACC, 2018 illustrating an expanded role of pathogenic deep intronic splicing variants in the pathogenesis of HCM, the authors ought to consider alerting readers that this protocol, as written, would miss such variants. That said, expansion of this approach to a genome-wide approach may help identify intronic/intergenic "hotspots" that underlie monogenic disease.
- 3) The MAF cutoffs proposed in this protocol are reasonable for the assessment of highly penetrant monogenic-disease causative variants. However, we are slowly learning that the genetic architecture of most cardiomyopathies and cardiac channelopathies, in many cases, is much more complex than the current monogenic/single gene model. As such, I would urge the authors to consider adding a sentence or two to the discussion that highlights for readers that so-called "risk alleles", variants with MAFs in gnomAD far greater than disease prevalence, will not be assessed by this approach. However, these variants may still contribute to cardiovascular diseases we still consider to be monogenic [this is illustrated in the channelopathy realm by p.Asp85Asn-KCNE1 (LQT5; Giudicessi et al, Circulation, 2018 and Lane et al, Heart Rhythm, 2018) and in the cardiomyopathies by p.Val122Ile-TTR/p.Val142Ile-TTR (Familial Amyloid Cardiomyopathy; Buxbaum and Ruberg, Genet Med, 2017) and p.del25-MYBPC3 (HCM; Dhandapan et al, Nat Genet, 2009)]. It is important to alert readers that the approach/MAF cut-offs outlined in this protocol will not assess these relatively common, but potentially clinically important variants.

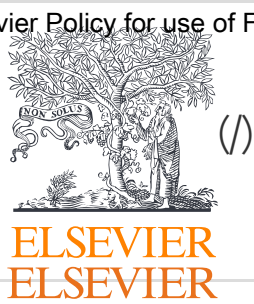
Both excellent points. Our protocol's signal-to-noise resolution is limited to exonic changes that are represented by amino acid changes, which largely omits intronic genetic variants. We now address this in the discussion as a method limitation, with reference to Bagnall et al's work. Further, we address the reviewers concern that our protocol may also exclude common potentially pathogenic variants by way of a MAF threshold, and note this as well as a limitation.

“As detailed, this protocol focuses specifically on amino-acid level changes resulting from exonic gene variants that code for amino acids, which excludes the role that pathogenic intronic splicing variants may play in monogenic disease. Given their recently demonstrated role in cardiomyopathies, expansion of the resolution this approach may be warranted to identify intergenic “hotspots” as well. Furthermore, the application of a MAF threshold may miss certain “risk alleles” that, though existing in the population with a MAF higher than that of disease prevalence, may contribute to disease

pathogenesis^{5,6}. Despite these limitations, this analysis is adaptable and can play a key role in providing clinicians a relative probability of disease pathogenicity when appropriate applied.”

Below are the citations for the above text changes. Please note that the numbering does not reflect the order of citations in the revised manuscript

- 1 Landstrom, A. P. *et al.* Amino acid-level signal-to-noise analysis of incidentally identified variants in genes associated with long QT syndrome during pediatric whole exome sequencing reflects background genetic noise. *Heart Rhythm*. 15 (7), 1042-1050, (2018).
- 2 Wang, Q. *et al.* Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat Genet*. 12 (1), 17-23, (1996).
- 3 Kapa, S. *et al.* Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. *Circulation*. 120 (18), 1752-1760, (2009).
- 4 Ackerman, M. J. *et al.* Ethnic differences in cardiac potassium channel variants: implications for genetic susceptibility to sudden cardiac death and genetic testing for congenital long QT syndrome. *Mayo Clin Proc*. 78 (12), 1479-1487, (2003).
- 5 Bagnall, R. D. *et al.* Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. *J Am Coll Cardiol*. 72 (4), 419-429, (2018).
- 6 Giudicessi, J. R., Roden, D. M., Wilde, A. A. M. & Ackerman, M. J. Classification and Reporting of Potentially Proarrhythmic Common Genetic Variation in Long QT Syndrome Genetic Testing. *Circulation*. 137 (6), 619-630, (2018).



Personal use

Authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes as outlined below:

- Use by an author in the author's classroom teaching (including distribution of copies, paper or electronic)
- Distribution of copies (including through e-mail) to known research colleagues for their personal use (but not for Commercial Use)
- Inclusion in a thesis or dissertation (provided that this is not to be published commercially)
- Use in a subsequent compilation of the author's works
- Extending the Article to book-length form
- Preparation of other derivative works (but not for Commercial Use)
- Otherwise using or re-using portions or excerpts in other works

These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article. In all cases we require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

Solutions

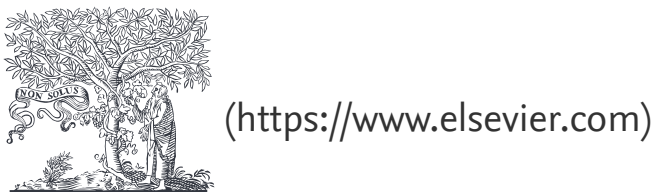


Solutions

Researchers



Researchers



ELSEVIER

Copyright © 2018 Elsevier, except certain content provided by third parties

Cookies are used by this site. To decline or learn more, visit our Cookies ([//www.elsevier.com/legal/use-of-cookies](https://www.elsevier.com/legal/use-of-cookies)) page.

[Terms and Conditions \(//www.elsevier.com/legal/elsevier-website-terms-and-conditions\)](https://www.elsevier.com/legal/elsevier-website-terms-and-conditions) [Privacy Policy \(//www.elsevier.com/legal/privacy-policy\)](https://www.elsevier.com/legal/privacy-policy) [Sitemap \(//www.elsevier.com/sitemap\)](https://www.elsevier.com/sitemap)



ELSEVIER

 RELX Group™ (<https://www.relx.com/>)

From: Permissions Helpdesk permissionshelpdesk@elsevier.com
Subject: RE: Permission for to use figure
Date: April 26, 2018 at 7:52 AM
To: Landstrom, Andrew landstro@bcm.edu



CAUTION: This email is not from a BCM Source. Only click links or open attachments you know are safe.

Dear Andrew,

Elsevier journal authors retain the right to use or re-use portions or excerpts in other works, subject to proper acknowledgement; see <https://www.elsevier.com/about/our-business/policies/copyright/personal-use>. As this is a retained right, no written permission is necessary provided that the material is not credited to any third party and that the original source is properly acknowledged.

If I may be of further assistance, please let me know.

Best Wishes,
Laura

Laura Stingelin

Permissions Helpdesk Associate

ELSEVIER | Global E-Operations Books

+1 215-239-3867 office

[l.stingelin@elsevier.c](mailto:l.stingelin@elsevier.com)

[om](mailto:l.stingelin@elsevier.com)

Contact the Permissions Helpdesk

+1 800-523-4069 x3808 | permissionshelpdesk@elsevier.com

From: Landstrom, Andrew [<mailto:landstro@bcm.edu>]
Sent: Wednesday, April 25, 2018 6:48 PM
To: Permissions Helpdesk <permissionshelpdesk@elsevier.com>
Subject: Permission for to use figure

*** External email: use caution ***

Dear Elsevier,

I am writing to ask permission to utilize a figure, published in the following article, in the supplement. I am the first and corresponding author. The article is:

"Amino acid-level signal-to-noise analysis of incidentally identified variants in genes associated with long QT syndrome during pediatric whole exome sequencing reflects background genetic noise." Heart Rhythm 2018 Mar 2. pii: S1547-5271(18)30142-5. doi: 10.1016/j.hrthm.2018.02.031. [Epub ahead of print]

PMID: 29501670

The figure I would like to use is supplemental figure 2 which is a figure of Kv7.1

I would like to include it as an “example” of amino acid-level pathogenicity analysis for a methods paper that will be submitted to the Journal of Visual Experimentation (JOVE).

The figure will come with a citation for the article and will state “Reprinted with Permission” in the figure legend.

Happy to answer any questions.

Andrew

Andrew Landstrom, MD, PhD
Section of Cardiology, Department of Pediatrics
Baylor College of Medicine

Texas Children's Hospital
1102 Bates Avenue
Feigin Center, Suite 430.09
Houston, TX 77030

Clinic: (832) 826-5600
Office: (832) 824-4122
Fax: (832) 825-0237
[Email: landstro@bcm.edu](mailto:landstro@bcm.edu)