InVivo Biosystems

Bridging the ACDY5 Diagnostic Gap

A Methodology Discussion

Anna Malinkevich and Chris Hopkins

Bridging the ACDY5 Diagnostic Gap

A Methodology Discussion

Anna Malinkevich and Chris Hopkins

Abstract:

Variants of Uncertain Significance (VUS) represent a startling percent of the reported alleles in disease causing genes, yet the functional consequence of VUS remains unclear. Uncertainty within the variant pool drastically reduces the utility of diagnostic genetic tests, inhibits clinicians from obtaining an unfettered understanding of patients' disease biology, and obstructs the possible pursuit of suitable therapies. The InVivo Biosystems Platform addresses this unique challenge by explicitly targeting the ambiguous variant landscape via humanized animal models.

The InVivo Biosystems Platform methodology creates humanized animal models to serve as Clinical Avatars of an individual patient or patient population. *C. elegans* humanized *ADCY5* model strains can rapidly probe a patient's disease biology, provide distinct functional analysis of the pathogenic verses benign variant landscape, and yield high quality assessments of VUS pathogenicity. This innovative technique provides the tools for rapid analysis of genomic variants towards the ultimate goals of generating risk stratification models, increasing patient management options, and potentially repurposing existing therapeutics.

Not All Variants are Created Equal: Variants of Uncertain Significance and the Diagnostic Gap

One in 15 people suffer from a rare disease. The reason something so seemingly common can still be considered rare is that the cause of such disease is distributed among 7,000 disease-associated genes, together comprising an estimated 30% of the annotated genome.[8,9,17,19] Accumulating genomic data reveals that each person has approximately 1,000 sequence variants that result in ambiguous, non-benign missense or indel mutations specifically in the coding regions of their genome – most compelling is that many of these variants are deemed as Variants of Uncertain Significance (VUS) or otherwise classified as having "unknown" disease repercussions.[9] With one-third of the genome estimated to be involved in disease biology, any single individual is likely to harbor over 300 codon-changing variations in their important "disease" genes.[8] Frameshifting indels with a high likelihood of pathogenicity account for only 7% of these variants. The functional consequences of the remaining variants are less clear. As a result, there remains a significant number of enigmatic alleles that exist in the background of anyone's personal genome.

As large-scale genome sequencing is increasingly being used in the clinical setting, variant interpretation has become a major bottleneck in accessing and exploiting the full utility of patient genetic information. Both physicians and clinician-scientists now face the escalating challenge of determining whether a suspect allele is directly contributing to disease as a pathogenic variant, or is simply observed as a non-consequential benign variant. Currently, over 40% of the individual genetic differences detected in a patient's genome remain unidentified in regard to pathogenic vs. benign state, and those alleles remain branded as VUS – Variants of Uncertain Significance.[3] The low diagnostic yield produced by clinical genomic sequencing and the growing diagnostic gap generated by the lack of variant classification continue to impede both efficient disease interpretation and optimized therapy strategies.[1]

ADCY5-related Dyskinesia: A Broad Variant Landscape Needing a Fast Animal Model

Per the ClinVar database, 60% of the reported variants of the enzyme adenylate cyclase 5 (ADCY5) are currently noted as VUS (Fig. 1).[3] This protein, encoded by the *ADCY5* gene, is responsible for the condition known as *ADCY5*-related Dyskinesia, which is a disorder characterized by abnormal involuntary movements.[7] To date, at least 400 cases of this disorder have been diagnosed and at least six individual *ADCY5* gene mutations are reported to be directly causative of the disease, due to an enzymatic gain-of-function (GOF) mechanism. The disorder is primarily diagnosed via genetic testing and, while *ADCY5*-related Dyskinesia is presently understood to be driven by autosomal-dominant mutations, it is not yet clear how known alternate loss-of-function (LOF) alleles ultimately contribute to disease biology.

Functional data derived from model system studies holds significant promise in revealing the details of *ADCY5* allele pathogenicity and reducing the uncertainty clouding reported *ADCY5* VUS alleles. Functional studies *in vivo* are highly influential for variant assessment, and variant function testing in rodent models is a well-established, gold-standard approach.[5,6,21] However, generation of accurate rodent models is both slow and exceedingly costly, highlighting the fact that rodent-derived data therefore

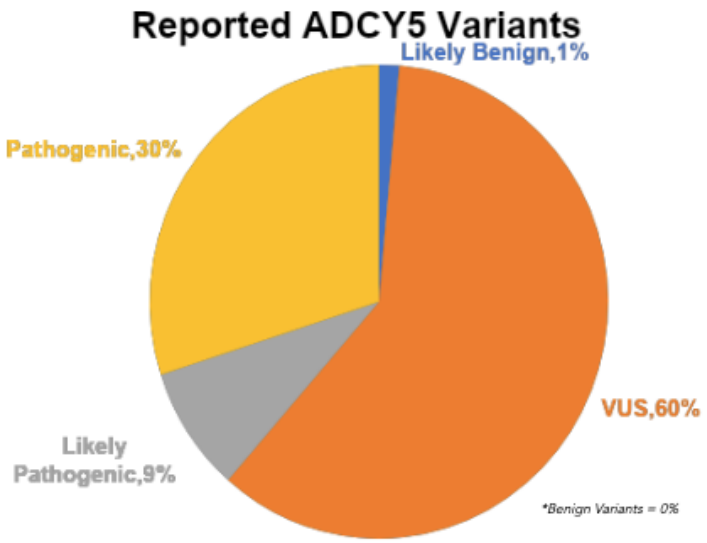

Figure 1. Human ADCY5 variant pool classification breakdown, as reported in ClinVar October 2019. *ClinVar variant classification is based upon the 5-tier classification system standards and guidelines recommended by the American College of Medical Genetics and Genomics (AMCG).*[28]

cannot be obtained within clinically relevant timescales nor rapidly reintegrated with a patient's original genetic report.

On the opposite end of the spectrum, technologies from bioinformatics to biochemical and cellular assays can be deployed to assess the functional consequence of a variant change. Induced pluripotent stem cells (iPSCs) and other cell culture methods have risen in relevance, but they fall short in addressing the physiological effects that relate to a patient's large-scale whole-body development or even to small-scale extracellular interactions.[4] Deep mutational scanning approaches for variant profiling have also been used in yeast; however, only 47% of examined human disease genes exhibited cDNA complementation in the yeast system.[10,26]

Conversely, the nematode *C. elegans* presents a model system than can rapidly and inexpensively interrogate human VUS on a whole-organism scale by providing both a high degree of human disease gene orthology and the experimental speed necessary for direct clinical relevance. In the case of *ADCY5*, the *C. elegans* ortholog *acy-4* exhibits 53% sequence identity to the human gene – granting a solid base of homology upon which to engineer a patient-specific humanized *C. elegans* animal model. Emergent CRISPR-based technologies allow for rapid, low-cost genetic manipulations of the nematode, which can then be deployed in a variety of biochemical and behavioral assays necessary to detect subtleties in variant biology on a whole-organism scale.

The InVivo Biosystems Platform: Modeling and Analysis of Humanized *ADCY5* Variants

The InVivo Biosystems Methodology

The InVivo Biosystems Platform is a revolutionary, patient-centric disease modeling approach leveraging the speed and tractability of the *C. elegans* nematode to generate humanized experimental animal models that accurately recapitulate a patient's allelic signature. Our proprietary, NextGen Humanization method

replaces the endogenous *C. elegans* ortholog with the human coding sequence for *ADCY5*, modified as needed to represent the allele variants in question. To establish an appropriately controlled – and therefore most experimentally informative – analysis environment, humanized *ADCY5* positive and negative control animals are generated along with the variant strains. Positive and negative controls are run through a series of benchmarking tests, collectively referred to as Strain Utility Assays, and the results subject to stringent cutoff thresholds to ensure that the engineered humanized models most adequately mimic human gene function. Only when the set quantitative metrics of the Strain Utility Assays have been successfully achieved by the control animals, do the *ADCY5* humanized variant strains undergo a battery of analyses towards ultimate classification of variant pathogenicity. A Training Set of known and confirmed pathogenic and benign *ADCY5* variants installed into humanized nematodes is functionally analyzed, and machine learning techniques are utilized to mathematically define clusters of pathogenic and benign variant classes. Humanized *ADCY5* VUS animals are then interrogated using these same parameters, and mathematically qualified as a best-fit for pathogenic or benign clusters within the variant landscape. Humanized variant strains, whether found pathogenic or benign, can then further be used for downstream investigation of compound efficacy and possible therapeutic strategies.

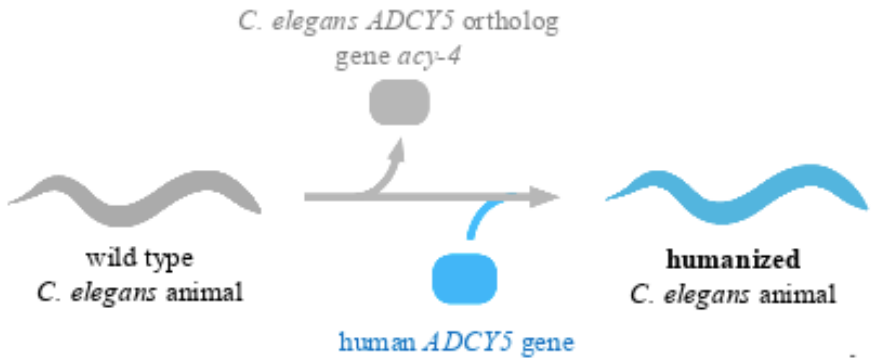

Figure 2. Scheme for the generation of C. elegans humanized animals to serve as Clinical Avatars for the interrogation of VUS allele pathogenicity.

NextGen Humanization

Humanization of an endogenous animal locus centers on two key attributes during design and engineering – codon bias and splicing elements. In order to adapt the GC-rich elements of the human genome for proper function in the AT-rich genomic environment of *C. elegans*, codon optimization is performed to capitalize on common nematode codon usage within the human gene.[20,25] As introns are integral for proper gene expression and regulation, splicing elements and at least three synthetic introns are then introduced into the codon-optimized human sequence.[11] This process is routinely performed in our labs in an average of 6 weeks, with an internal competition resulting in the generation, installation and functional testing of a clinical variant in humanized strains in under 10 days.

Negative Control

In genetic studies, a true negative control requires a deletion of the entire coding sequence in question. Complete functional loss or deletion of *C. elegans ADCY5* ortholog *acy-4* has not yet been fully investigated; it is possible that *acy-4* is an essential gene, loss or removal of which may create lethality in the nematode. A partial deletion of *acy-4* is currently classified as lethal, but remains to be further supported by experimental evidence.[27] Interestingly, in mice, deletion of the *ADCY5* locus does not

have neonatal lethality.[18] In order to obtain the most credible and conclusive experimental data possible, a full knock-out (KO) of *C. elegans ADCY5* ortholog *acy-4* is generated to serve as an absolute null of the genetic locus under manipulation. Should loss of *acy-4* result in lethality, the deletion strain animals are maintained as heterozygous stocks.

Positive Controls

Introduction of a fully-functioning, canonically 'normal' human *ADCY5* gene into the endogenous *C. elegans acy-4* locus will, conceptually, generate a positive control strain. However, a positive control remains merely conceptual until this wild-type *ADCY5* humanized strain has shown to successfully rescue nematode activity back to *C. elegans* wild-type levels – ie: when the human *ADCY5* gene functionally mimics endogenous *acy-4* activity in the worm, establishing and/or confirming a baseline level of normal function in the humanized animal. Functional benchmarking analysis of positive controls is performed as a series of Strain Utility Assays (described below), which serve to quantitatively evaluate the degree of rescue in humanized vs. wild type nematodes.

To further control for any possible artifacts introduced by gene recoding and the transgenic humanization process itself, a supplementary process control animal is simultaneously created. *C. elegans acy-4* endogenous genetic material is manipulated and processed the same way as the human *ADCY5* gene, and then inserted back into the nematode as a recoded transgene. If the recoding, sequence optimization and gene insertion procedure render a gene-swapped animal that mimics the biochemical and behavioral profile of the wild type nematode, this is indicative of minimal or no genetic artifacts introduced via the genetic engineering process itself. With a solid negative control, verified positive control, and supplementary process control, the stage is properly set for initial benchmarking followed by introduction of humanized variants and functional analysis.

Benchmarking

Prior to the humanization of variant strains and final interrogation of the experimental set via functional analysis, control animals are subject to a series of Strain Utility Assays in order to confirm a desirable utility profile. This serves to identify and optimize any control strains exhibiting deleterious or undesirable background mutations, and to establish conforming molecular expression profiles among all controls. A 10% variation threshold is maintained for all of the molecular and biochemical assays performed – animals must exhibit a Strain Utility Assay response within 10% of that measured in wild type *C. elegans* in order to proceed forward into final variant testing analysis.

Training Sets and the Variant Landscape

In order to accurately define the variant landscape, a Training Set of known and confirmed pathogenic and benign human *ADCY5* variants are generated within humanized animals; this group of pre-defined variants is characterized for functional activity via Gene Function Assays, and qualified into clusters using machine learning techniques. A conformance threshold of 90% must be met in order proceed to final VUS analysis – at minimum, 90% of known benign and 90% of known pathogenic variants must produce functional data consistent with their assignment as reported in the literature.

VUS Analysis

Upon successful benchmarking and acquisition of a mathematically defined variant landscape, patient *ADCY5* VUS variant alleles are integrated into humanized animal models and investigated for pathogenicity. A Variant Assessment Engine is used to incorporate animal model data from a patient VUS variant and compare its activity directly with the established Training Set. Clustering algorithms are used to bin VUS activity as consistent with either pathogenic or benign behavior, and binning is enhanced by "human understandable" machine learning approaches.[2] Machine learning outputs in classifying previously-VUS variants as pathogenic or benign are collected and the data delivered as a Variant Function Report. These now-classified variants can be further interrogated to answer questions regarding therapeutic compound efficacy and to plan possible therapeutic strategies.

Targeted VUS Analysis: The InVivo Biosystems Platform Bridges the Diagnostic Gap

Variants of Uncertain Significance represent a startling percent of the reported alleles in disease causing genes, yet the functional consequence of VUS remains unclear. Uncertainty within the variant pool drastically reduces the utility of diagnostic genetic tests, inhibits clinicians from obtaining an unfettered understanding of patients' disease biology, and obstructs the possible pursuit of suitable therapies. The InVivo Biosystems Platform addresses this unique challenge by explicitly targeting the ambiguous variant landscape via humanized animal models.

The InVivo Biosystems Platform methodology creates humanized animal models to serve as Clinical Avatars of an individual patient or patient population. *C. elegans* humanized *ADCY5* model strains can rapidly probe a patient's disease biology, provide distinct functional analysis of the pathogenic v. benign variant landscape, and yield high quality assessments of VUS pathogenicity. This innovative technique provides the tools for rapid analysis of genomic variants towards the ultimate goals of generating risk stratification models, increasing patient management options, and potentially repurposing existing therapeutics.

References

1. Baldridge D, et al. The Exome Clinic and the role of medical genetics expertise in the interpretation of exome sequencing results. Genet Med. 2017 Sep; 19(9): 1040-1048. https://www.ncbi.nlm.nih.gov/pubmed/28252636

2. Boulesteix AL, et al. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. WIREs Data Min Knowl. 2.6 (2012): 493-507.

3. NCBI ClinVar. https://www.ncbi.nlm.nih.gov/clinvar

4. Csöbönyeiová M, et al. Recent advances in iPSC technologies involving cardiovascular and neurodegenerative disease modeling. Gen Physiol Biophys. 2016 Jan; 35(1): 1–12. https://www.ncbi.nlm.nih.gov/pubmed/26492069

5. Cox TC. Utility and Limitations of Animal Models for the Functional Validation of Human Sequence Variants. Mol Genet Genomic Med. 2015 Sep; 3(5): 375–82. https://www.ncbi.nlm.nih.gov/pubmed/26436102

6. Eilbeck K, et al. Settling the score: variant prioritization and Mendelian disease. Nat Rev Genet. 2017 Oct; 18(10): 599- 612. https://www.ncbi.nlm.nih.gov/pubmed/28804138

7. Genetics Home Reference https://ghr.nlm.nih.gov/

8. Hegde M, et al. Development and validation of clinical whole-exome and whole-genome sequencing for detection of germline variants in inherited disease. Arch Pathol Lab Med. 2017 Jun; 141(6): 798-805. https://www.ncbi.nlm.nih.gov/pubmed/28362156

9. Jansen IE, et al. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. Genome Biol. 2017 Jan 30; 18(1): 22. https://www.ncbi.nlm.nih.gov/pubmed/28137300

10. Kachroo AH, et al. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science. 2015 May 22; 348(6237): 921-5. https://www.ncbi.nlm.nih.gov/pubmed/25999509

11. Lacy-Hulbert A, et al. Interruption of coding sequences by heterologous introns can enhance the functional expression of recombinant genes. Gene Ther. 2001 Apr; 8(8): 649-53. https://www.ncbi.nlm.nih.gov/pubmed/1132041212.

References

12. Lemke JR, et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia. 2012 Aug; 53(8): 1387-98. https://www.ncbi.nlm.nih.gov/pubmed/2261225713.

13. Li Q and Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet. 2017 Feb 2; 100(2): 267-280. https://www.ncbi.nlm.nih.gov/pubmed/2813268814.

14. Liang JS, et al. Genetic diagnosis in children with epilepsy and developmental delay/mental retardation using targeted gene panel analysis. Neuropsych. 2018 Oct; 8(5): 1577-1585.15.

15. Liu J, et al. Novel and de novo mutations in pediatric refractory epilepsy. Mol Brain. 2018 Sep 5; 11(1):48 https://www.ncbi.nlm.nih.gov/pubmed/3018523516.

16. Ma C, et al. Reconstitution of the vital functions of Munc18 and Munc13 in neurotransmitter release. Science. 2013 Jan 25; 339(6118): 421-5. https://www.ncbi.nlm.nih.gov/pubmed/23258414

17. Mallick S, et al. The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations. Nature. 2016 October 13; 7624(538): 201-6. https://www.ncbi.nlm.nih.gov/pubmed/27654912

18. MGI – Mouse Genome Informatics http://www.informatics.jax.org/

19. Morgan T, et al. Non-validation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. JAMA. 2007 April 7; 297(14): 1551-61. https://www.ncbi.nlm.nih.gov/pubmed/17426274

20. Piovesan A, et al. Universal tight correlation of codon bias and pool of RNA codons (codonome): The genome is optimized to allow any distribution of gene expression values in the transcriptome from bacteria to humans. Genomics. 2013 May; 101(5): 282-9. https://www.ncbi.nlm.nih.gov/pubmed/23466472

21. Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May; 17(5): 405-24. https://www.ncbi.nlm.nih.gov/pubmed/25741868

22. Saitsu H, et al. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. Nat Genet. 2008 Jun; 40(6): 782-8. https://www.ncbi.nlm.nih.gov/pubmed/18469812

References

23. Saitsu H, et al. STXBP1 mutations in early infantile epileptic encephalopathy with suppression-burst pattern. Epilepsia. 2010 Dec; 51(12): 2397-405.
https://www.ncbi.nlm.nih.gov/pubmed/20887364

24. Stamberger H, et al. STXBP1 encephalopathy: a neurodevelopmental disorder including epilepsy. Neurology. 2016 Mar 8; 86(10): 954-62.
https://www.ncbi.nlm.nih.gov/pubmed/26865513

25. Sun X, et al. An improved implementation of effective number of codons (nc). Mol Biol Evol. 2013 Jan; 30(1): 191-6.
https://www.ncbi.nlm.nih.gov/pubmed/22915832

26. Weile J, et al. A framework for exhaustively mapping functional missense variants. Mol Syst Biol. 2017 Dec 21; 13(12): 957.
https://www.ncbi.nlm.nih.gov/pubmed/29269382

27. WormBase
https://www.wormbase.org/

28. Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May; 17(5): 405-24.
https://www.ncbi.nlm.nih.gov/pubmed/25741868