DECIPHER Workshop

Julia Foreman, EMBL-EBI

Introduction

This tutorial will provide you with training in using DECIPHER to classify variants and find similar patients. By the end of this workshop participants will be:

- familiar with DECIPHER's main variant classification interfaces
- able to classify sequence variants using DECIPHER
- able to classify copy number variants using DECIPHER

You do not need to login to complete this exercise. If you have a login, please log out.

Part 1: Analysis of Sequence Variants

Case History – Sally

Sally is a three year old female (46XX) with global developmental delay, short stature, elbow hypertrichosis, epicanthus, microcephaly and feeding difficulties. She is the second child of healthy, unrelated parents. You have referred Sally for Whole Genome Sequencing. Samples from her parents were also sent for analysis. The lab report details two candidate diagnostic variants (GRCh38):

- Heterozygous *de novo* variant in *CDK13* NM_003718.5:c.3185T>C
- Heterozygous de novo variant in *KMT2A* NM_001197104.2:c.3556A>T

Sally's DECIPHER record

https://www.deciphergenomics.org/patient/519521

A patient page in DECIPHER is divided into sections, each containing specific bits of deposited or derived information. We will review the:

- Overview
- Phenotype tab
- Genotype tab

The genotype page is divided into two main sections. The top section provides information about the variants. Clicking on the variant location of any variant activates the bottom section of the page which contains tabs which relate to the selected variant.

Variant Classification: The genome browser

The browser is centered on the variant showing that our patient's variant is a T>C change. The genomic location of our patient's variant is shaded in grey in all tracks to help orientate the user. By default the genome browser tracks shown include: Genes, This Patient: Sequence Variants, DECIPHER CNVs, DECIPHER Sequence Variants. The user can select which tracks they would like to be displayed in the browser by clicking on the "Tracks" button on the top left of the browser.

Genome Browser Track descriptions

Chevrons "<<" on the right side of tracks indicate additional filters and further information.

- ***This Patient: Sequence Variants:*** This patients' sequence variant(s).
- ***Genes track.*** Gene(s) that are within the variant.

    - *By default, the gene(s) are coloured according to the probability of haploinsufficiency (pHaplo; i.e., deletion intolerance). Genes can also be coloured by the predictive scores pLI, LOEUF or SHet.*
- ***Morbid Genes:*** Genes associated with disease phenotypes curated by OMIM.
- ***Conservation:*** Conservation of protein sequence between species for ortholog genes.
- ***Missense Constraint:*** Genes/gene regions coloured by missense depletion.
- ***Transcripts:*** This track shows all transcripts.
- ***Gene Disorder variants:*** This track displays variants associated with known disorders as annotated by GeneReviews.
- ***Population dataset tracks:***
    - ***gnomAD Sequence Variants:*** Sequence variant data from the Genome
    - ***gnomAD Structural Variants:*** Structural variant data from the Genome
    - ***gnomAD Coverage:*** gnomAD sequence coverage information.
- ***Disease dataset tracks:***
    - ***DECIPHER: Sequence Variants***: Sequence variants observed in other DECIPHER patients.
    - ***ClinVar Sequence Variants:*** Variant data from the ClinVar database.
    - ***HGMD Public:*** Sequence variants reported in the public version of the Human Gene Mutation Database (HGMD).
    - ***LSDB Variants:*** Variants from the Locus-specific databases (LSDB) maintained by the Leiden Open Variation Database (LOVD).
- ***Regulatory features:*** Features from the Ensembl Regulatory Build, which are predicted to regulate gene expression.

1. **What is the genomic location of the *CDK13* gene?**
   Answer:
2. **Is the CDK13 gene predicted to be haploinsufficient according to the pHaplo score?**
   Answer:

   Is *CDK13* known to be associated with developmental disorders?
   A. **Is CDK13 associated with a developmental disorder according to the Gene2phenotype list?**
      Answer:
   B. **If *CDK13* is associated with a developmental disorder according to the Gene2phenotype list, what is the disorder? What is the mode and consequence?**
      Answer:
   C. **Is the CDK13 gene a Morbid Gene and if so which disease phenotypes is it associated with?**
      Answer:
3. **Our patients has a T>C change, which causes a Valine to Alanine amino acid substitution. Is this Valine well conserved across different species?**
   Answer:
4. **Is this region of the *CDK13* gene missense constrained?**
   Answer:
5. **Are there any variants in the same location reported in gnomAD (population database)? Is it the same variant?**
   Answer:
6. **What is the total allele frequency of this *CDK13* variant in gnomAD according to genome sequencing?**

Answer:

7. **Are there any variants in the same location reported in disease databases (ClinVar, HGMD, DECIPHER)? If so, what is their reported pathogenicity?**
   Answer:

Variant Classification: The protein browser

The protein view shows a 2D protein structure with the important domains of the protein highlighted, such as the protein kinase domain in CDK13. The yellow parallel line indicates the location of Sally's CDK13 variant.

Variants from **disease datasets** are shown directly above (DECIPHER) and below (ClinVar).
Loss of function variants are shown in red and protein changing (missense) in yellow. Variants annotated as likely/definitely pathogenic are displayed as closed triangles/squares, all other variants are shown as open triangles/squares. Protein truncating codons are shown as squares.
Information from the **population dataset** gnomAD is shown towards the top of the protein browser. These tracks include the location of missense variants, homozygous missense variants, loss of function (LoF) variants and homozygous LoF variants.
Missense constraint data is shown above the gnomAD variant data. A predicted **NMD (non-sense mediated decay) escape** region is shown towards the top of the view and **Exon** structure is shown at the top of the protein browser.
If **3D structures** of the protein are available, these are shown at the bottom of the protein browser. Experimentally determined structures are displayed, in addition to AlphaFold predicted structures. It is also possible to view a **3D structure** of the protein (or part of the protein) by clicking on the protein structure in the 3D structure track.

If you use your mouse wheel to zoom into a region, an **amino acid sequence** track will also be displayed. It is possible to customise the protein browser to display only your preferred tracks and to apply filters by clicking on the "Tracks and filters" button above the protein browser.

8. **Is Sally's CDK13 variant in the same protein domain as other reported likely pathogenic/ pathogenic DECIPHER and ClinVar variants?**
   Answer:

Variant Classification: VEP

For every sequence variant deposited in DECIPHER, the Ensembl Variant Effect Predictor (VEP) is run which predicts consequences for the variant across all transcripts.
ClinGen have a recent publication suggesting thresholds for CADD and REVEL (https://pubmed.ncbi.nlm.nih.gov/36413997)

- CADD scores ≥25.3 indicate evidence for pathogenicity and score ≤22.7 benignity
- REVEL scores ≥0.644 indicate evidence for pathogenicity and score ≤0.290 benignity.

9. **Do the predictive scores suggest the variant is deleterious to gene function/product?**
   Answer**:**

Variant Classification: Matching patients

The matching patient variant interface shows phenotype information about patients with sequence variants in your gene of interest. At the top of the page are summary statistics about

patients with variants in *CDK13*. There are also summary lists about how well your patient's phenotype matches other patients with variants in *CDK13*. Below the summary statistics is a table showing patients in DECIPHER that have sequence variants in *CDK13*.

10. **Sally's CDK13 variant is protein changing. What percentage of matching patients are known to have protein changing variants?**
    Answer:
11. **Sally has six phenotypes. Are there any shared phenotypes between Sally and other patients with <u>likely pathogenic/ pathogenic</u> CDK13 variants?**
    Answer:
12. **Which of Sally's phenotypes most often overlaps with other patients with <u>likely pathogenic/ pathogenic</u> *CDK13* sequence variants? How many overlapping patients have this phenotype?**
    Answer:
13. **Which of Sally's phenotypes does not overlap with other patients with likely pathogenic/ pathogenic *CDK13* sequence variants?**
    Answer:

Variant Classification: Gene tab

At the top of the page is information about the gene and about known gene/disease associations. At the bottom of the page is aggregated quantitative phenotypic data.

14. **Are there any phenotypes shown here where patients harbouring CDK13 variants have a matching phenotype to Sally's? If so, which phenotype(s)?**
    Answer:

15. **Which ACMG criteria apply and is this a pathogenic variant?**
    Answer:

Recording Pathogenicity – Demonstration (login required)

We will now assess the pathogenicity of the variant using the sequence variant pathogenicity module in DECIPHER.

Variant Classification: Independent work

Using the same approach, analyse the *KMT2A* variant reported for this patient and answer the following questions.

16. **Is the *KMT2A* gene predicted to be haploinsufficient according to the pHaplo score?**
    Answer:
17. **Is *KMT2A* known to be associated with developmental disorders?**
    A. **Is KMT2A associated with a developmental disorder according to the Gene2phenotype list?** This is a list of gene associations curated by UK consultant clinical geneticists and is shown by clicking on the *KMT2A* gene in the Genes Track.
    Answer:
    B. **If *KMT2A* is associated with a developmental disorder according to the Gene2phenotype list, what is the disorder? What is the mode and consequence?**
    Hint: click on the KMT2A gene in the Genes track

Answer:

C. **Is the KMT2A gene a Morbid Gene and if so which disease phenotypes is it associated with?** Morbid genes are displayed in the Morbid Genes track. Associated disease phenotypes are shown by clicking on the *KMT2A* gene in the Morbid Genes track.
Answer:

18. **Are there any variants in the same location reported in gnomAD (population database)? Is it the same variant?**
Answer:
19. **Is there gnomAD exome and genome sequence coverage in this region of the *KMT2A* gene?**
Answer:
20. **Are there any variants in the same location reported in disease databases (ClinVar, HGMD, DECIPHER)? If so, is it a loss of function variant?**
Answer:
21. **Is the KMT2A variant in any specific domain in the protein?**
Answer:
22. **Are there many reported likely pathogenic/ pathogenic DECIPHER and ClinVar loss of function variants in the KMT2A?**
Answer:
23. **What is the predicted consequence of Sally's KMT2A variant? Does the CADD score suggest it is extremely deleterious?**
Answer:
24. **Sally's KMT2A variant is loss of function. What percentage of matching patients are known to have loss of function sequence variants?**
Answer:
25. **Sally has six phenotypes. Are there any shared phenotypes between Sally and other patients with <u>likely pathogenic/ pathogenic</u> KMT2A variants?**
Answer:
26. **Which of Sally's phenotypes most often overlaps with other patients with likely pathogenic/ pathogenic KMT2A sequence variants? How many overlapping patients have this phenotype?**
Answer:
27. **On the Gene (clinical) tab, are there any phenotypes shown here where patients harbouring KMT2A variants have a matching phenotype to Sally's? If so, which phenotype(s)?**
Answer:
28. **Which ACMG criteria apply and is this a pathogenic variant?**
Answer:

Recording Pathogenicity – Demonstration (login required)

We will now assess the pathogenicity of the variants using the sequence variant pathogenesis module in DECIPHER.

DECIPHER
GRCh38

Summative assessment (Case Interpretation) - Demonstration

DECIPHER has a Summative Assessment tool to support multidisciplinary teams working to evaluate whether a variant(s) explains the clinical features seen in a patient. This tool uses a framework to evaluate the 'Clinical fit' and then enables the user to select the OMIM gene-disease pair that best fits the patient and to select possible options to summarize the fit.

The summative assessment tool is based on following evidence framework:

	Evidence against genotype-phenotype relationship		**Evidence for genotype-phenotype relationship**			
	Strong	**Supportive**	**Supporting/ weak**	**Moderate**	**Strong**	**Very strong**
Genomic footprint of phenotype in the genome – specifically of clinical features seen in the patient for a given gene	Phenotype is very distinctive and not caused by this gene	Phenotype is non-specific and not caused by this gene	Phenotype is non-specific and can be caused by 31-100 genes	Phenotype is not very specific and can be caused by 11-30 genes	Phenotype can be caused by variants in 3-10 genes	Very distinctive phenotype associated with one or two genes
Age at presentation (onset of symptoms) matches typical representation for condition caused by this gene/variant	Inconsistent with what is currently known for this genetic disorder	Poor match	Consistent with what is currently known for this genetic disorder	Excellent match		
Clinical fit – how closely does this patient match a typical patient for this disorder in terms of features assessed by clinical history, examination or measurement.	Characteristics almost always seen in this genetic disorder are absent in the patient or characteristics almost never seen in this disorder caused by this gene/variant are present (and not explained by a second diagnosis)	Characteristics commonly seen in this genetic disorder are absent in this patient	Consistent with the genetic diagnosis in question	Typical of the genetic diagnosis in question	Very good match for the diagnosis in question	Excellent match for the diagnosis in question
Severity and progression of clinical features and signs	Very atypical and uncharacteristic of the genetic diagnosis in question	Not usually seen in the genetic diagnosis in question	Typical of the genetic diagnosis in question			
Additional clinical investigations e.g. imaging, biochemistry, metabolic investigations, histology	Results of specific investigations are almost never seen in the genetic disorder in question e.g. normal cranial MRI in LIS1	Results of specific investigations are not typical for the genetic disorder in question	Results of specific investigations are consistent with in the genetic disorder in question e.g. normal cranial MRI in child with FRAX	Results of specific investigations are suggestive of the genetic disorder in question e.g. mild ventricular hypertrophy in HCM	Results of specific investigations are diagnostic for gene/variant in question but could also be seen in a few similar conditions e.g. multiple adenomas in Lynch syndrome	Results of specific investigations are pathognomonic for gene/variant in question (or are seen in only a very few genetic conditions e.g. Kayser-Fleisher rings in Wilson's disease or reduced PAH activity in PKU or positive sweat test in CF)
Relevant family history (based on clinical reporting of family history)	Irreconcilable with the gene/variant(s) and mechanism identified	Not consistent with the gene/variant(s) and mechanism identified	Consistent with the gene/variant(s) and mechanism identified	Substantial family history Consistent with the gene/variant(s) and mechanism identified	Extensive family history Consistent with the gene/variant(s) and mechanism identified	

The Summative assessment tool in DECIPHER can be accessed by selecting the "Assessments" tab. A new assessment is created by clicking on the "New assessment" button.

Logged out mode: Search

Variant pages e.g. NM_000257.4:c.1208G>A, NM_001101426.4:c.1251G>A
Protein pages: PACS1:p.R203W

Part 2: Analysis of CNVs

Case history – Davina

Davina is an eight year old with tall stature, macrocephaly, global developmental delay, mild intellectual disability, constipation. Davina also displays autistic behaviour and anxiety. She is the second child of healthy, unrelated parents. WGS revealed that Davina has a 2.1Mb *de novo* deletion on chromosome 14 (46XX).

Davina's DECIPHER record

https://www.deciphergenomics.org/patient/519524

Variant Classification: Dosage sensitivity scores

Dosage sensitivity scores are based on the predicted haploinsufficiency of individual genes within the deletion or duplication. A high positive score indicates that the CNV is more likely to be pathogenic than a CNV with a lower positive score, or a negative score. The dosage sensitivity sampling probability is an estimate of the proportion of the general population that carry a rare deletion/duplication with a dosage sensitivity score as severe or more severe than for the CNV being assessed.

Variant Classification: The Genes Table

This displays a list of genes that overlap the CNV with gene predictive scores and curated gene-disease information. The location column indicates if the CNV overlaps the whole gene or part of the gene.

29. **How many genes does the CNV overlap in total? How many of these genes are protein coding?**
    Answer:
30. **How many genes are listed if we use the "Likely dose sensitive genes" filter at the top of the table?**
    Answer:
31. **Are there any genes that are predicted to be very haploinsufficient (probability >0.9) using the pHaplo score?**
    Answer:
32. **Does Davina's CNV overlap any genes that have been curated by ClinGen as being haploinsufficient? If so, which gene(s)?**
    Answer:
33. **Does Davina's CNV overlap any genes for which G2P has reported the variant consequence associated with a disease as absent gene product? If so, which gene(s)?**
    Answer:
34. **Does Davina's CNV overlap all of this known haploinsufficient gene?**
    Answer:

Clicking on a row in the table activates the bottom section of the page which contains a clinical and protein/genomic tab which relate to that gene.

Variant Classification: The Genome Browser

The genome browser for CNVs, by default, has some of the same tracks as sequence variants and some additional tracks.

35. **Are there any similar likely pathogenic/pathogenic losses in DECIPHER or ClinVar?**
    Answer:
36. **Are there any similar losses in population datasets (*gnomAD SV, DGV gold, Population: CNV)*?**
    Answer:
37. **Does the CNV encompass any ClinGen DS regions or DECIPHER CNV syndromes?**
    Answer:

Variant Interpretation: Matching Patients

This will display the matching patient interface for CNVs. Since Davina has a deletion the loss filter is applied by default. The interface is similar to that displayed for sequence variants. We can also use the information on the sequence variant tab.

38. **How many of Davina's phenotypes match patients with pathogenic/likely pathogenic, likely LOF CHD8 sequence variants?**
    Answer:
39. **What percentage these sequence variants are *de novo*?**
    Answer:

Recording Pathogenicity – CNV loss

We will now assess the pathogenicity of the variants using the ACMG/ClinGen CNV loss interface in DECIPHER. If the CNV is small and it is relevant to use the sequence variant guidelines, there is an option to use this framework instead.

A further demonstration patient can be found here:
https://www.deciphergenomics.org/patient/519967/

Further information

DECIPHER: Improving Genetic Diagnosis Through Dynamic Integration of Genomic and Clinical Data: https://pubmed.ncbi.nlm.nih.gov/37285546/
DECIPHER: Supporting the interpretation and sharing of rare disease phenotype-linked variant data to advance diagnosis and research: https://pubmed.ncbi.nlm.nih.gov/35143074/