TEACHING OUTREACH RESOURCE

Rare Disease Genetics

Brought to you by

Genetics
Mātai Ira
otago.ac.nz/genetics

genetics Otago

University of Otago
ŌTĀKOU WHAKAIHU WAKA

GO BEYOND WITH GENETICS

__Explore__ the origins and mechanics of life. __Discover__ causes of disease and their cures. __Solve__ the problems facing our agriculture and natural heritage. __Understand__ the past. __Create__ a better future. __Master__ the world of genetics...

WELCOME...

…to this Genetics teaching resource, created by Genetics Otago and the Genetics Teaching Programme at the University of Otago.

Our aim is to engage young minds with Genetics and to do this we have developed a range of resources that include information, worksheets and activities or experiments that will help you to plan exciting Genetics classes for your students.

Where possible we have endeavoured to align and link the content of the resources to the New Zealand Curriculum.

If you have any questions relating to the content of the resources or would like to organise an onsite teaching session on one of our topics please contact us at go@otago.ac.nz.

Contents

Welcome ii
Contents iii

Overview 1

Part A
- Rare Diseases 4
- Genetic Variation 5

Part B
- Introduction to Sequence Analysis 10
- Ethical and Cultural Considerations 12

Part C
- Family A 15
- Analysis of Genome Data 16
- Gel Electrophoresis of *HIVEP2* 19

Worksheets

Overview

Rare diseases are individually rare, but collectively common, affecting 3.5-5.9% of the global population or around 300,000 people in Aotearoa. Many of these diseases have a genetic basis and the majority of these are unknown. In this resource, we provide background information, instructions and worksheets necessary to provide a basic understanding of the genetics of rare diseases. Students will be given the opportunity to work through a series of exercises to understand the concepts before identifying the genetic cause of a rare disease and testing for it in a family group. Please be aware that the story and results in this resource are based on a real family, so please ensure that the exercises and surrounding discussion be completed with respect.

OBJECTIVES

- Successfully follow a scenario through a logical set of steps to reach an informed conclusion
- Understand how rare diseases arise and the bioinformatic and molecular methods used to identify the cause.
- Appreciate that the search for genetic variants is difficult, time-consuming and not always conclusive; and appreciate that answers can be meaningful beyond a diagnosis for patients and their families.
- Interpret the results of an electrophoresis gel and apply the results to the scenario.

SECTIONS

Part A: Rare Diseases
- Rare Diseases Overview
- Genetic Variation
    - Activity One – **Identify the Mutations**
    - Activity Two – **Old or New**

Part B: Bioinformatics
- Introduction to Sequence Analysis
- Ethical and Cultural Considerations
    - Activity Three – Ethics discussion/debate

Part C: Case Study

- Family A
    - Activity Four – **Family A Diagnosis (part A)**
- Analysis of Genome Data
    - Activity Five – **Family A Diagnosis (part B)**, using '**Family A Genome Data**'
    - Activity Six – **Family A Diagnosis (part C)**
- Gel Electrophoresis of HIVEP2
    - Activity Seven – **Family A Diagnosis (part D)** (can be completed with a practical component up-on request)

CIRRICULUM LINKS

This resource is designed to feed into the following curriculum areas at level 7+.

- **Nature of Science**
    - *Understanding about science* – connecting ideas and knowledge to historical knowledge and making it available to future investigators.
    - *Investigating in science* – carrying out complex investigations and understanding the relationship between these investigations
    - *Communicating in science* – Using appropriate vocabulary and conventions during the investigation and reporting of results
    - *Participating and contributing* – develop an understanding of the socio-scientific issues and identify responses at personal and societal levels.

- **Living World**
    - *Life processes, ecology and evolution* – Explore the effects of genetic variation and appreciate their impact biologically, socially and ethically.

ACHIEVEMENT STANDARDS

This resource is relevant to the following standards:

- 2.5/ 91157 Demonstrate understanding of genetic variation and change
- 2.7/ 91159 Demonstrate understanding of gene expression
- 3.7/ 91607 Demonstrate understanding of human manipulations of genetic transfer and its biological implications *(as an introduction to the processes)*

PART A

Rare Diseases

Rare Diseases

A rare disease is one that affects less than 1 in 2000 people, there are currently more than 7,000 such diseases identified globally, and the number continues to grow. This means that although each individual disease is rare, collectively they are relatively common with an estimated 5-10% of individuals (in other words up to 800 million people) affected by a rare disease in 2022.

While not all rare diseases have a genetic origin, the vast majority (~80%) do and as our understanding of genetics and genomics grows so too does our understanding of this class of disease. Currently, our understanding places rare diseases into one of four categories:

1. Single Gene – caused by single, or multiple variants in one gene
2. Chromosomal – caused by a change in the structure or number of chromosomes
3. Multifactorial – caused by a combination of genetic and environmental factors
4. Non-genetic – believed to have no genetic cause (these can include types of cancer, autoimmune diseases and infections)

As alluded to by the categories of rare diseases variants causing rare diseases can be as small as a single nucleotide variant (SNV) and as large as a whole chromosome. They can be inherited from parents or occur spontaneously (de novo) in the germline or developing embryo. Due to the complex nature of the genetics and the small number of patients with any one disease, it is difficult to see the causative smaller variants in a patient above the noise of normal human variation. Over the last 20 years, a lot of progress has been made in diagnosing rare genetic diseases but many patients and their families still remain either undiagnosed or with no information about the cause of their disease. As our ability to collect and analyse full genome sequences has become better, faster, and cheaper so too has our ability to identify small variations causing disease. An important part of this, that is still lacking for the majority of minority populations, is having an appropriate reference genome to compare patients to.

Rare diseases often present with a wide range of symptoms that often differ to some extent between patients, making it more difficult to identify groups of patients with the same condition. While we are making gains in identifying the cause of some of these diseases upward of 90% still remain without a treatment let alone a cure.

Genetic Variation

Genetic variation (or mutation) can arise from two types of changes; chromosomal changes or relatively smaller sequence changes.

Chromosomal changes refer to a change in the number or structure of the chromosomes. These changes can occur by the following means:

- Deletion: the deletion of a large portion of or an entire chromosome. Examples include Turner syndrome – monosomy X i.e. loss of one sex chromosome and Cri-du-chat syndrome - Deletion of the end of the short arm of chromosome 5.
- Duplication: the duplication of a large portion of or an entire chromosome. An example of duplication is Down syndrome – trisomy 21 i.e. 3 copies of chromosome 21.
- Translocation: the change in the position of a portion of a chromosome to either another region within the same chromosome (intrachromosomal) or to another chromosome entirely. Translocations can be reciprocal i.e. a broken portion of one chromosome is swapped with a portion of a second non-homologous chromosome, or non-reciprocal i.e. a chromosomal segment moves from one location to another without the exchange of a second segment. The Philadelphia chromosome is an example of translocation between the long arms of chromosomes 22 and 9 which produces a gene fusion leading to leukaemia.
- Inversion: The change in orientation of a segment of chromosome i.e. the chromosome breaks in two places and the segment rotates 180^{o} before reattaching into the same chromosome. An inversion can be paracentric (does not involve the centromeric region) or pericentric (does involve the centromeric region). Disruption of the factor VIII gene due to inversion has been shown as a cause of haemophilia.

Sequence changes originate at the DNA level and are shown at the protein level (DNA -> RNA -> protein) and are typically divided into point mutations and frameshift mutations.

Point mutations are the substitution of one base for another (e.g. an A becomes a C in the DNA sequence), these can be silent (no change to the amino acid sequence or protein) missense (alters the sequence in a way that changes a single amino acid within the protein) or nonsense (introduces a stop codon prematurely truncating the protein).

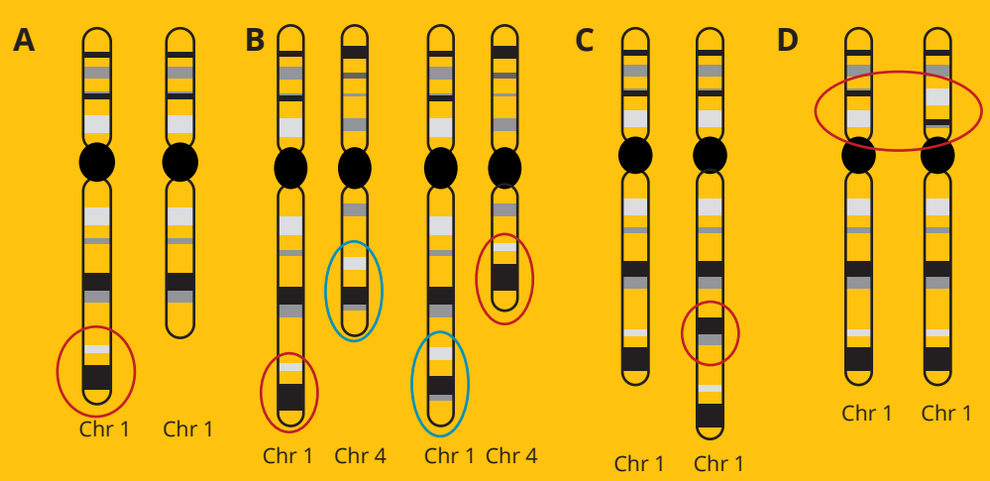
Frameshift mutations occur due to an insertion or deletion of one or more bases in the DNA sequence. Because the translation of RNA (transcribed from DNA) occurs in triplets this causes the reading frame to change having a dramatic effect on the

downstream amino acid sequence and hence protein. Frameshift mutations can be missense or nonsense and almost always result in a non-functional protein.

All types of variation can be either inherited (passed from a parent to a child) or de novo (occurring for the first time in a child due to a variant in a germ cell (egg or sperm) of one parent, or in the fertilized egg itself).

Activity One: Have the students complete the worksheet **'Identify the Mutations'**, they will need to be provided with access to the Amino acid codon table (there is a copy of this in the back of this manual).

ANSWERS



E *AUG CUU GAG UGG GCU AGG AUG ACG UGG CUA G*
Met Leu Glu Trp Ala Met Thr Trp Leu X
This is an **insertion** *of a C between bases 20 and 21 of the original sequence causing a* **frameshift mutation**.

F *AUG CUU GAG UAG*
Met Leu Glu Stop
This is a C->T **substitution** *at position 11 in the original sequence causing a* **nonsense mutation**.

G AUG CUU GAG UGG GCU AGG AUC CGU CCC UAG
Met Leu Glu Trp Ala Arg Ile Arg Pro Stop
This is a T->G **substitution** *at position 21 in the original sequence causing a* **silent mutation**.

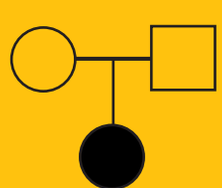
H AUG CUU GAG UGG ACU AGG AUA CGU CCC UAG
Met Leu Glu Trp Thr Arg Ile Arg Pro Stop
This is a C->T **substitution** *at position 13 in the original sequence causing a* **missense mutation**.

- *F is likely to be the most pathogenic mutation as it causes a severely truncated protein.*
- *E is also highly likely to be pathogenic as it has changed the reading frame meaning that the scond half of the protein is now different to the original (this also has the follow on effect that there will be extra amino acids incorporated into the protein until an intact stop codon is reached).*
- *H may be pathogenic depending on how important the changed amino acid is to protein structure (how it folds) and if it (the amino acid) is important for binding of other proteins.*
- *G is highly unlikely to be pathogenic as it does not change the protein at all.*

Activity Two: Have the students complete the worksheet **'Old or New'**.

ANSWERS

A



De novo mutation, future children have no increased risk over general population.

B

```
Reference:TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC

Patient:  TAC GAA CTC ACC CGA TCC TGC AGG GAT C
          TAC GAA CTC ACC CGA TCC TGC AGG GAT C
Parent 1: TAC GAA CTC ACC CGA TCC TGC AGG GAT C
          TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Parent 2: TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
          TAC GAA CTC ACC CGA TCC TGC AGG GAT C
```

Deletion of bases A and T from positions 20 and 21 of the sequence (c.20delAT) resulting in a frameshift mutation. The variant is inherited from both parents meaning that PKD is autosomal recessive. The parents show no symptoms as a single copy of the intact gene produces sufficient functional protein. Any future children of this couple have a 25% change of inheriting both copies of the variant and therefore having PKD.

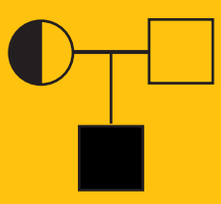
C

```
Reference:TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC

Patient:  TAC GAA CTC ATC CGA TCC TAT GCA GGG ATC
          TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Parent 1: TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
          TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Parent 2: TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
          TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Sibling 1:TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
          TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Sibling 2:TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
          TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
```

The variant is a substitution of a C for a T at position 11 (c.11C>T) resulting in a Trp amino acid becoming a stop codon (p.W4X). This is nonsense substitution (resulting in a truncated protien). The variant is not present in either parent so is *de novo*.

D



The variant is inherited from the mother.

There will be no difference in the risk in either case. The donor is very unlikely to have Haemophilia A or he would have known and not been allowed to make the donation. There is a 50% chance that any future children will inherit the variant from the mother. If the child is XX they will be a carrier like the mother and if the child is XY they will have the disease (becasue the disease is X linked and a single functional copy of the gene is protective).

PART B

Bioinformatics

Introduction to Sequence Analysis

The first entire human genome was published in 2003 and the advances in genetic medicine since then have been exponential with continued advances in technology. Now, 20 years later, it is reasonable to sequence the entire genome of not only patients with diagnosed or suspected rare diseases but also their parents to gain insight into not only the possible cause of the disease but also the origin of it.

As sequences are added to databases around the world the power of bioinformatics grows and so too does our ability to identify causative variants. This means that individuals with undiagnosed disorders or atypical phenotypes can undergo this testing and have all of their genetic variations considered for candidacy as the causative genetic factor for their illness.

With the amount of natural variation segregating in human populations, a large number of variants will be found for an individual. But in the case of a Mendelian disease, only one of these variants (or possibly two for a recessive disorder) can be the cause. The remainder will be benign, or at the very least unrelated to the disease under consideration.

For the purposes of this module, we are going to focus on whole exome datasets (and later apply the learning in a clinical scenario). Because the whole exome is the search space it is useful to consider the search for the mutation as the application of a set of filters. No filter is perfect, and no filter can be applied without making some assumptions about where the mutation might be and what alteration it might encode. For instance, the decision to perform an exome analysis in the first place assumes:

- That the mutation is inside the exome (true for ~85% of Mendelian disorders);
- That the mutation is a sequence-level anomaly (i.e. not a copy number variant or epigenetic problem)

Some key points to remember are:

1. Each scenario is unique and therefore the search strategy differs for each.
2. For some presentations the “search space” i.e., the number of genes

that need to be considered differs. For instance, for some classical presentations, there can be extraordinary locus heterogeneity and any one of dozens of genes could conceivably house the mutation

3. For some disorders, the type of mutation differs. While for most a loss of function mechanism applies some can have gains or alterations in function
4. Key to this search strategy is the use of on-line databases of human genetic variation.
    - Ostensibly healthy individuals (beware, these individuals like all of us, will still be heterozygous for many deleterious recessive alleles)
    - Disease-specific databases. Catalogues, organised by disease type or locus that list variants causative of a given disease state (beware that the level of evidence of causation can vary enormously within and between such databases)
5. Missense variants can be difficult to arrive at definitive conclusions about whether they are pathogenic or not. Short of performing functional studies, some attributes of these variants can contribute to forming a judgement. Typically, pathogenic missense variants that lead to rare monogenic highly penetrant disorders are:
    - Rare in catalogues of genetic variation in humans
    - Substitute phylogenetically conserved residues
    - Disruptive of functional or structural domains of proteins

A variety of tools are available where the genetic coordinates of a missense variant can be entered and a score produced that gives an indicative guide to pathogenicity. These programmes (Polyphen2, SIFT, mutation taster etc) are around 75-80% sensitive in assigning pathogenic status to missense genetic variants.

USEFUL WEBSITES

Polyphen2: http://genetics.bwh.harvard.edu/pph2/
SIFT: https://sift.bii.a-star.edu.sg/
Mutation Taster: https://www.mutationtaster.org/
Online Mendelian Inheritance in Man® (OMIM®): https://omim.org/
ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
National Center for Biotechnology Information (NCBI): https://www.ncbi.nlm.nih.gov/

Ethical and Cultural Considerations

At its core ethics is about values, and to act in an ethical manner researchers must ensure that any study reflects the values held by the individual(s) or group(s) involved e.g. upholding tikanga for Māori patients.

DNA is a gift, and it is a taonga not only in its form as blood or tissue but also as data. Due to this, careful consideration must be given to the collection, storage and use of the DNA and data as well as the return of the biological material (or a representation of it, such as reports or information on the results). In the same way that a patient must give informed consent for a medical procedure, informed consent must be gained for genomic research and due to the nature of genomic research projects this informed consent must be maintained for the life of the project, particularly where the findings lead to an evolution of the aims.

It is also important to remember that genomics will inform on other members of a patient's family by proxy, even if they are not included in the research. This has ethical implications particularly if a close relative of a patient does not want to know about their disease status. This is a key consideration for Māori who, in order to protect wairua, must ensure that all whanau are comfortable with the involvement of a patient in the research.

While researchers are obligated to do their best to uphold the values of the participants in the ways stated above, it is not always possible to account for all eventualities. Incidental findings are one such eventuality in research using genomic information and consideration must be given to the appropriate course of action if such findings are made. Ethics are not always black and white.

Activity Three: Break the class into small groups and give each group a scenario to discuss in terms of ethical implications. Some example scenarios:

- Should all babies have their genomes sequenced at birth and if so who should have access to the data?
- A child and their father decide to have their DNA sequenced by an online ancestry company, and the child realises from the results that they are not related to their father. Should the son tell his 'father'?

- Should health insurance companies have access to genomic data?
- A 15-year-old whose grandfather has a late-onset dominant heritable disease wants to be tested for the gene mutation. The child's mother (daughter of the grandfather) does not want to know her disease status. Should the child be able to have the test?
- What are the implications of cheap and accessible direct to consumer genome sequencing?

PART C

Case Study

Family A

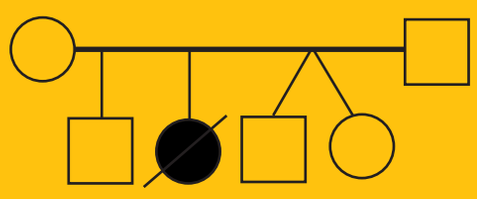
Family A has a young son and baby daughter, both children were well at birth and the son continues to develop as expected. At six weeks of age, the daughter had an MRI of her brain in response to an episode of sleep apnea which showed a brain consistent with other infants of her age, with no abnormalities. No further episodes of apnea occurred, and the baby continued to grow as expected. However, by six months old the baby was showing signs of developmental delay such as being unable to roll over or lift her head. At seven months baby A was admitted to the hospital with a severe respiratory infection where she began having what appeared to be seizures. These were not febrile convulsions and an EEG showed that there was lessened brain activity during the episodes. The baby girl was given another MRI of her brain which showed significant brain damage caused by atrophy to both the white and grey matter that was not localised and was progressive. There was no evidence of trauma, stroke or infection of the brain. The result left the infant with a 6–12-month prognosis, but the baby tragically died less than 6 weeks later. Family A has since welcomed twins, a girl and a boy, who are now 2 years old and are healthy and developing normally.

The parents and baby girl had whole genome sequencing done as a part of the diagnostic process, as it was deemed that the child likely had a rare genetic disease. The remainder of this module will focus on this data.

Activity Four: Have the students complete part A of the worksheet '**Family A Diagnosis**'.

ANSWERS

A



Above is the pedigree with all of the information we are aware of. It is most likely that this is a *de novo* mutation in Baby A as none of the other children show symptoms of the disease. However we can't rule out that the parents are carriers of the variant making it an autosomal recessive disease. This could be the case if, by chance, none of the remaining 3 children recieved the variant from **both** parents

Analysis of Genome Data

RESOURCES

If you wish to do this activity with your students please get in touch (go@otago.ac.nz) to request access to the data file and a video tutorial for teachers of howwork through this activity with your students.

Activity Five: The students will now complete part B of the '**Family A Diagnosis**' worksheet by analysing the genome data in file '**Family A Genome Data**' to identify candidate genes. Students will need to be given access to the file and should work in 6 small groups from this point on. Provide the students with the following instructions, or walk through the exercise together explaining the steps as you go.

Inheritance Model Filter:

In the genotypes columns set the genotypes that model the inheritance pattern you would like to filter on. Here we can test the two most likely models autosomal recessive, where the parents would be heterozygous and the child homozygous for the alternate allele, and *de novo,* where the alternate allele is only found in the child. A *de novo* mutation should only affect one allele (a *de novo* mutation affecting both alleles is incredibly unlikely).

0/0 = homozygous for the reference allele
0/1 **and** 1/0 = heterozygous for the alternate allele
1/1 = homozygous for the reference allele

Remember you are setting the genotype to find the variant i.e. what you expect the genotype of the parents and the daughter (Baby A) to be.

Coding effect filter:

Your list of variants is still very long, and it would be nice to prune it down to those that are most likely to disrupt gene function. In the ***Impact*** column, un-check all except "HIGH". This means you are only keeping the variants that change the sequence of the protein in some way and are therefore likely to have a pathogenic effect.

Allele frequency filter:
Our disease is rare, and as a consequence the allele frequency of any variant that causes the disease is likely to be low in most populations. For example if a recessive disease affects 1 in 1,000,000 then the allele frequency of any single variant causing the disease is likely to be less than 1 in 1000 (0.001). Let's remove all the variants that are too common to cause a rare disease. The header to filter on here is ***Population allele frequency*** this is the highest allele frequency found for the variant in any of the populations shown to the right.

In the filter menu under 'Number Filters', select the 'Less than or equal to' option. Then to the right, type 0.01 in the box (i.e. we are allowing an allele frequency as high as 1 in 100, consistent with a disease prevalence of up to 1 in 400). That is a very conservative estimate, so we should be safe!

ANSWERS

B Following the above instructions and using a recessive inheritance model should give you 1 variant in the HLA-DRB5 gene. The students should see however that this variant has been reported previously to be benign so is very unlikely to be the cause of the disease.

Following the above instructions and using the *de novo* model should give you 12 variants in 9 different genes:

1. *NBPF14*
2. *PRR23A*
3. *MUC4*
4. *HIVEP2*
5. *VPS13B*
6. *C9orf3*
7. *WDFY4*
8. *KRTAP4-7*
9. *SCAF1*

Activity Six: Have the groups complete part C of the worksheet '**Family A Diagnosis**'.

Here they will research the known functions and any documented association with diseases of each of the 10 genes. This should further narrow their candidates to one or two genes. They may start by simply Googling the Gene names but below are some useful websites they might try (there are many others):

- Online Mendelian Inheritance in Man® (OMIM®): https://omim.org/
- ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/
- MedlinePlus: https://medlineplus.gov/
- Clinical Genome Resources (ClinGen): https://search.clinicalgenome.org
- Decipher Genomics: https://www.deciphergenomics.org
- GeneCards - The Human Gene Database: https://www.genecards.org

ANSWERS

C *HIVEP2* is the best fit based on the bioinformatics analysis and in terms of the phenotype relating to the known functions of this gene. Students should find mentions of intellectual disability, delayed development of motor skills and weak muscle tone. It is also known that one copy of the altered gene is sufficient to cause disease and that most cases are *de novo*.

Students may also suggest *NBPF14*, this variant is a possibility from the bioinformatics exercise as it has no data in clinVar or Decipher (rather than being ruled out by this data as is the case for *VPS13B*) A Google search of *NBPF14* should reveal that it too is linked to developmental and neurogenetic disease, it is however also linked to congenital heart and kidney disease, urinary tract abnormalities and cancer, symptoms Baby A dis not display, making it a less likely candidate.

There is no additional information to add to the pedigree because the genomic data show that neither parent has the variant (so they cannot be changed to carriers on the pedigree). This confirms that the variant is *de novo*.

Gel Electrophoresis of *HIVEP2*

From the information we have the most likely candidate for Baby A's rare disease is HIVEP2. Although this is shown to be a de novo mutation through the sequence analysis the family are anxious that their twins may have the same variant. To ease their concerns their doctor agrees to have the twins tested for the HIVEP2 mutation found in their sister. Rather than sequencing the entire genome, the lab can simply amplify the affected region via allele-specific polymerase chain reaction (PCR) which can then be examined via gel electrophoresis.

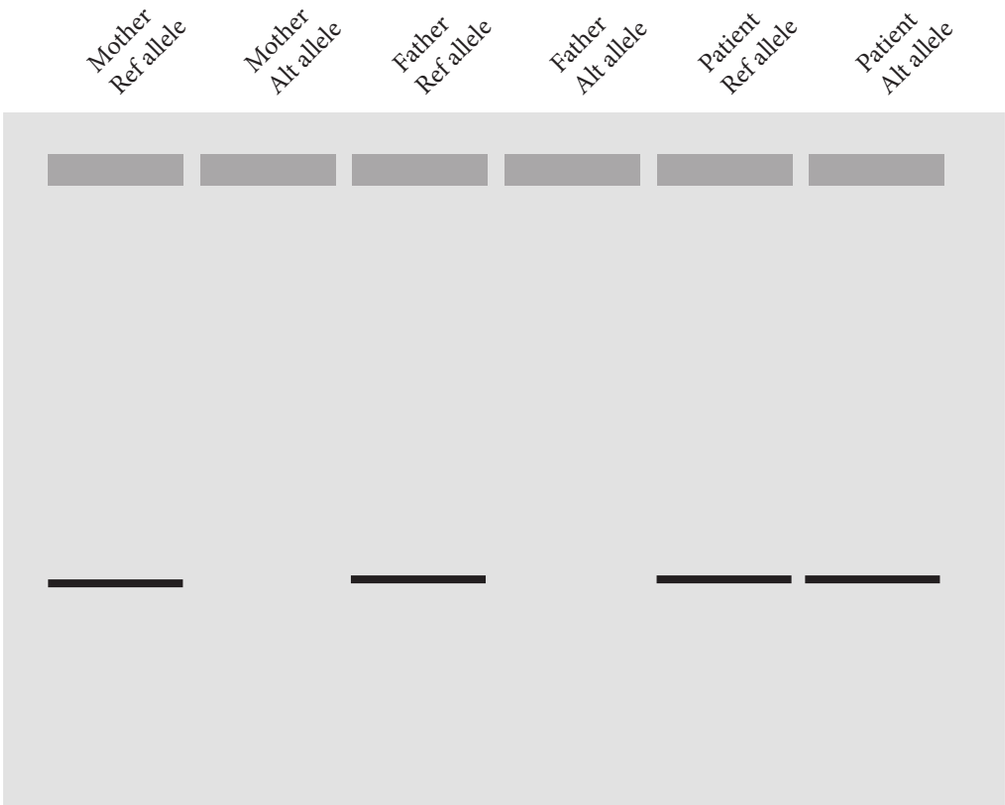
An allele-specific PCR uses two different sets of primers to identify the presence of the reference allele and/or the alternate (variant) allele. The reactions are carried out in two separate tubes and then each is run in its own lane of the electrophoresis gel. For peace of mind, the whole family are tested at this site.

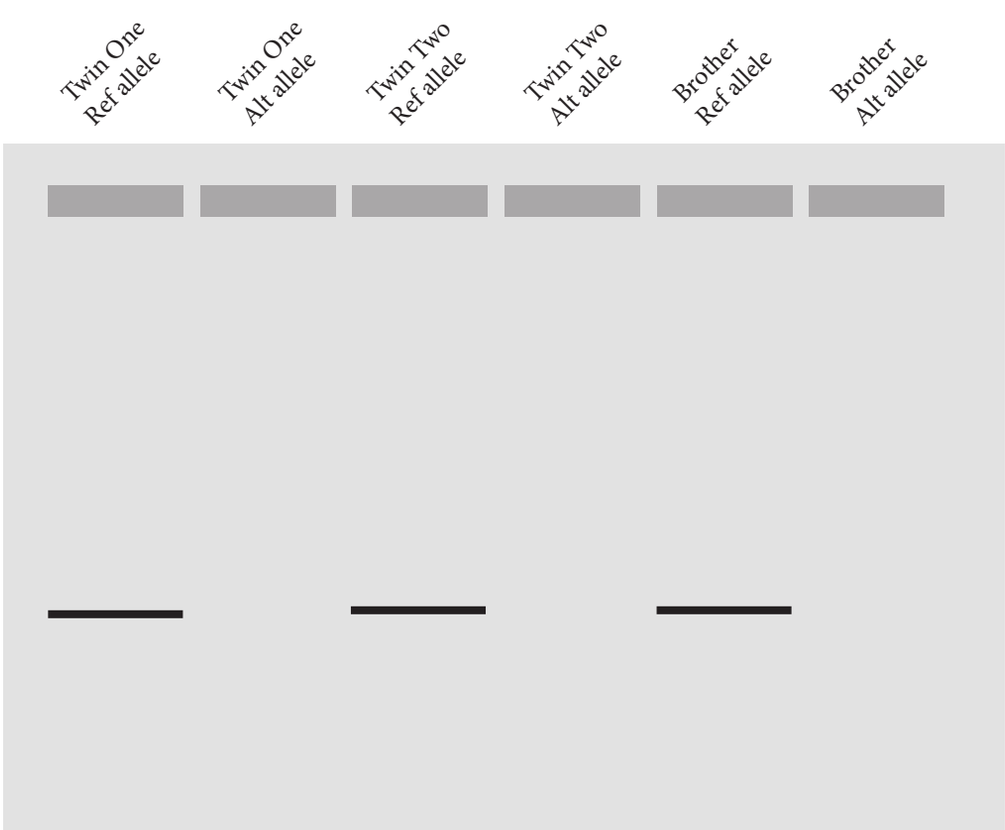
Activity Seven: Draw the gel diagrams on the next page on the board (or print copies for the students) before having the students answer part D of the '**Family A Diagnosis**' worksheet.

DO NOT do this if you plan to run the practical experiment, the students will use the results from the experiment to complete the worksheet.

ELECTROPHORESIS KITS

This final identification step can be done as a practical exercise using the Genetics on the GO kits available from us. In this case the students will make and pour an agarose gel that they will then load 'DNA' samples into and analyse the results to decide which individuals are affected by the disease and which are carriers. For this exercise we supply dye samples as the DNA to eliminate the need for staining of DNA meaning that the exercise can be done in a single lesson. If you would like to borrow one of these kits please contact us at go@otago.ac.nz.


Mother Ref allele
Mother Alt allele
Father Ref allele
Father Alt allele
Patient Ref allele
Patient Alt allele


Twin One Ref allele
Twin One Alt allele
Twin Two Ref allele
Twin Two Alt allele
Brother Ref allele
Brother Alt allele

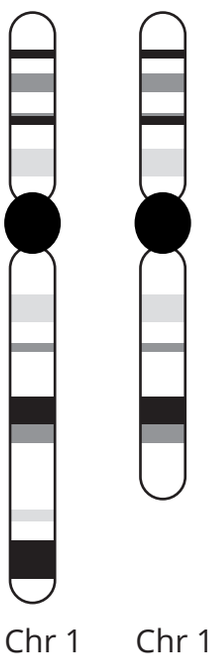
ANSWERS

D The banding pattern for the gels is shown on page 20. This shows that the patient is the only member of the family to carry the alternate allele (variant), meaning that no other members of the family are affected, and none can pass the variant on to their future children. The variant was truly *de novo* and is no more likely to recur in this family than in any random individual.

Identify the Mutation

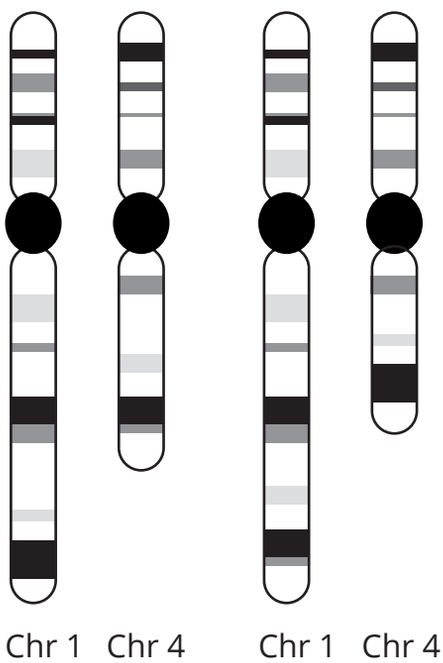
For each of the below examples circle the mutation present and identify the type of mutation.

A



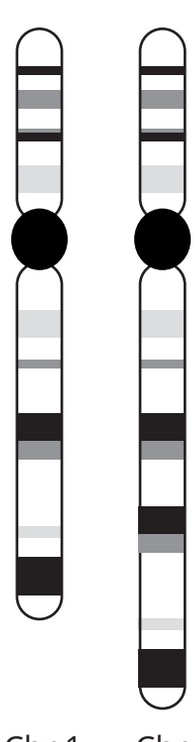
_____

B



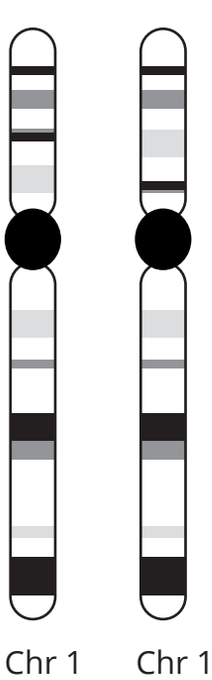
_____

C



_____

D



_____

Original DNA Sequence:	TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
mRNA Sequence:	AUG CUU GAG UGG GCU AGG AUA CGU CCC UAG
Amino Acid Sequence:	Met Leu Glu Trp Ala Arg Ile Arg Pro Stop

E Mutated DNA Sequence: TAC GAA CTC ACC CGA TCC TAC TGC AGG GAT C

mRNA Sequence: _____

Amino Acid Sequence: _____

Type of Mutation: _____

F Mutated DNA Sequence: TAC GAA CTC ATC CGA TCC TAT GCA GGG ATC

mRNA Sequence: _____

Amino Acid Sequence: _____

Type of Mutation: _____

G Mutated DNA Sequence: TAC GAA CTC ACC CGA TCC TAG GCA GGG ATC

mRNA Sequence: _____

Amino Acid Sequence: _____

Type of Mutation: _____

H Mutated DNA Sequence: TAC GAA CTC ACC TGA TCC TAT GCA GGG ATC

mRNA Sequence: _____

Amino Acid Sequence: _____

Type of Mutation: _____

Order E - H based on likely pathogenicity and explain your decision.

_____

_____

_____

_____

Old or New?

Rare diseases can be caused by inherited or *de novo* (new) mutations. In the below examples answer the questions to identify if the mutation is inherited or de novo and what this means for the families.

A Sarah and Mateo have a 1-year-old daughter who has been diagnosed with Achondroplasia. Achondroplasia is an abnormality of bone growth resulting in short stature among other skeletal differences. Sarah and Mateo are both within normal height ranges.

Draw a pedigree for the family

Is the mutation inherited or *de novo*? And what does this mean for future children of this couple?

B Two healthy parents have recently delivered a baby girl with Polycystic kidney disease (PKD). Sequencing of a gene known to cause PKD in other individuals was undertaken for this family, with the following results:

```
Reference:      TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC

Patient:        TAC GAA CTC ACC CGA TCC TGC AGG GAT C
                TAC GAA CTC ACC CGA TCC TGC AGG GAT C
Parent 1:       TAC GAA CTC ACC CGA TCC TGC AGG GAT C
                TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Parent 2:       TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
                TAC GAA CTC ACC CGA TCC TGC AGG GAT C
```

Circle the variant(s) in the sequence above. What type of mutation is this?

Why is the Child affected but not the parents?

Are future children likely to have PKD?

C A child was born with an unknown rare disease. Genome sequencing of the parents and two siblings was undertaken. Below is a short section of the sequence showing the likely causative variant.

```
Reference:      TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC

Patient:        TAC GAA CTC ATC CGA TCC TAT GCA GGG ATC
                TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Parent 1:       TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
                TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Parent 2:       TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
                TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Sibling 1:      TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
                TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
Sibling 2:      TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
                TAC GAA CTC ACC CGA TCC TAT GCA GGG ATC
```

Circle the variant(s) in the sequence above. What type of mutation is this?

Is the mutation inherited or *de novo*?

D Amaia conceived a child with the use of an anonymous sperm donor. The child, a boy, was born with Haemophilia A, an X-linked disorder. Subsequent testing showed that Amaia was a carrier of this disorder.

Is the mutation inherited or *de novo*?

Draw a pedigree for the family

Are future children of Amaia with i) the same donor and ii) a different donor likely to be affected? Explain your answer.

Family A Diagnosis

Work through the below exercises to identify the probable causative gene for baby A's illness.

A Draw a pedigree for the family

From the pedigree what do you think the most likely mode(s) of inheritance (dominant, recessive, *de novo*) of this rare disease is (are)? Explain your answer.

B Complete the bioinformatic exercise using the genome data provided in file '**Family A Genome Data**', following the instructions provided by your teacher.

List the possible Gene candidates you identified:

C Research what is known about the genes you identified in part B (a Google search of the gene name is a good place to start). Make some brief notes on any genes that seem to be likely candidates:

Based on her symptoms, and your analysis, which is the best candidate gene to have caused Baby A's illness?

Add any additional information you have gained from the bioinformatics to your pedigree, if there is no additional information what does this mean?

D Draw the banding pattern from the electrophoresis gels. What does this confirm about the variant?

	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Koha

While these kits will remain free for schools in need, we ask that you please consider a koha of whatever your school can afford to help in the continued development and upkeep of these ever-popular kits. Each kit costs us $50 - $120, depending on class size and location. Our resources have been used by more than 1500 students, from Invercargill to Whangarei, annually over the past few years.

You can donate by simply visiting:
https://alumni.otago.ac.nz/donate/genetics-otago-on-the-go

Thank you!

Feedback

We hope you have enjoyed this resource.

Feedback is very welcome to:
go@otago.ac.nz