How to read a genetic report: examples from Phelan-McDermid syndrome (PMS)

"This overview is intended to make genetic reports easier to read for parents. It was written with parents of children with PMS in mind, but may be useful to others. It is not a replacement for consultation with geneticists or genetic counselors; instead it is written to help make talking to medical professionals easier and more productive."

-- Teresa M. Kohlenberg, MD

[Last updated 16 March 2021]

The latest version of this document can be found online here:
https://arm22q13.wordpress.com/how-to-read-a-genetic-report-examples-from-phelan-mcdermid-syndrome-pms/

Part 1: Understanding terms that appear in genetic reports

A: Chromosomes, genes, and the proteins they make

Chromosome
A chromosome is a long curled-up chain of DNA. DNA provides instructions for all of the parts and functions of our bodies. Humans have 46 chromosomes that come in pairs (these are called "autosomes" and have numbers 1 to 23) plus two sex chromosomes that are not in pairs, X and Y. Males have an X and a Y. Females have two X chromosomes.

Nucleic Acids ("nucleotides")
Nucleotides are the building blocks of DNA. They are hooked together in pairs making a long twisting chain (called a double helix). There are only four nucleotides in the human DNA code, usually written by the first letters of their names: A=adenine, C=cytosine, G=guanine, T=thymine. Every group of three nucleotides is called a "codon". The cell reads the gene in groups of three and translates each codon into the amino acid it corresponds to.

Gene
A gene is a sequence of nucleic acids (nucleotides) along DNA. The sequence is a code for how to make proteins. Proteins are built from amino acids. Every group of 3 nucleotides signals one particular amino acid, so it is crucial that a gene sequence is transcribed and used in groups of 3.

Amino Acids

Amino acids are the building blocks that get hooked together based upon the DNA code in a gene to make proteins. Proteins are the brick and mortar of the body: they make the cells and organs of the body. Special proteins (enzymes) manage the chemical reactions of the body. Most neurotransmitters and hormones are proteins. There are about 20 amino acids used to make proteins, each typically identified by a 3-letter abbreviation.

RNA

This overview focuses on the DNA code and the resulting proteins. However, there are several steps in between. DNA is in a special central zone of the cell called the "nucleus". The machinery for making proteins is outside the nucleus in a zone called the "cytosol". In order to copy the DNA instructions, move them out of the nucleus, and then use the instructions to make many copies of the protein, the cell uses RNA. The process is complex and involves several different types of RNA (tRNA, mRNA, rRNA). This overview does not go into the RNA steps.

B. Types of genetic tests

Karyotype

Karyotyping involves opening up a cell and spreading the actual chromosomes out on a slide to study under a microscope. The observer can count the number of chromosomes and note anything unusual in their shape. For example, karyotyping is the preferred method for detecting a ring chromosome, or an extra chromosome such as the extra copy of chromosome #21 which causes Down Syndrome, or a missing chromosome such as in Turner Syndrome. Although chromosomal microarrays (see below) can detect small changes not visible with karyotyping, karyotyping can demonstrate the *structure* of deletions, duplications, insertions, unbalanced translocations, ring chromosomes and other abnormalities that a microarray may miss.

In-situ Hybridization

Sometimes "ish" or "FISH" shows up in a genetic report. ".ish" stands for "in-situ hybridization". It is an older method for studying deletions and other rearrangements. It is sometimes still used as a definitive test for certain rearrangements (e.g. balanced translocation).

Chromosomal Microarray

Often abbreviated as "CMA", it is a sensitive test that looks for extra (duplicated) or missing (deleted) chromosome segments.

Sequencing

Sequencing is a technique that read the letters of the DNA code in order, letter by letter. This provides very detailed information about a gene or multiple genes.

Whole exome sequencing (WES)

Sequencing the most relevant segments of every gene of the DNA. This is a broad but detailed picture of all the genes on all the chromosomes.

Whole genome sequencing (WGS)

Sequencing every letter of an individual's DNA. This is more detailed than WES.

C. Variants and mutations

Variant

All humans have the same set of genes (about 20,000 different ones) in their DNA, but different people can have different versions of any given gene. The different versions of a gene are called "variants". Variants account for many differences in things like hair color or whether a person has freckles or how likely they are to be tall or short.

Common variant

A variant that is found commonly in the general population. Common variants tend to be benign (not harmful).

Pathogenic variant

A variant of a gene that is known to be associated with disease.

Benign variant

A variant of a gene that has not caused disease in the past.

Variant of Unknown Significance (VUS)

VUS is used when you have no information about a variant, or when you suspect the variant is causing a problem, but you don't yet have enough evidence to label it "pathogenic".

De novo variant

The variant was not found in the parents but is found in the child.

Mutation

A term sometimes used to indicate a change in the DNA sequence that can cause damage or disorder. That use of the term has been recently replaced with the term "pathogenic variant."

Mosaic, mosaicism
When some of the cells tested have one variant and other cells have a different variant. For example, half of the cells could have a common variant while the other half may have a pathogenic variant.

Germ-line variant
A variant that is found only the parent's sperm or eggs, but that is not present in the parent's other body (somatic) cells. Mutations in germ lines (sperm or eggs) explain how, on rare occasions, parents who test negative for a pathogenic variant can have multiple children with a pathogenic variant.

Copy number variant / copy number variation (CNV)
Duplicated or missing chromosomal segments are sometimes called copy number variants (CNVs). Any addition or subtraction to the DNA code can be called a CNV. There is no minimal or maximal size when talking about CNVs, although once a whole chromosome is duplicated or deleted, the terminology changes to chromosome number variation.

D. Phenotype and penetrance

Phenotype
The way the individual appears and functions.

Penetrance
How often a variant (or other genetic abnormality) actually causes a change in the phenotype such as a disease. A highly penetrant variant almost always leads to the disease/disorder.

Incomplete penetrance
A term for a variant (or other genetic abnormality) that causes disease only sometimes. This is another explanation for how, on rare occasions, parents may have the variant but not have the disorder, while their children do have the disorder.

Autosomal dominant
Humans have two copies of every gene except for the genes on the sex chromosomes, which are not paired. Autosomal dominant means that one pathogenic variant is sufficient to cause disease. Pathogenic *SHANK3* variants are generally autosomal dominant. (*SHANK3* is an important gene in PMS because a variant can be highly penetrant, pathogenic and autosomal dominant.)

Part 2: How to understand the technical language for a sequence variant

Let's start with this example, a small part of the *SHANK3* gene. The top row shows the nucleotides, listed in groups of three. The bottom row shows the amino acids, one amino acid below each codon. Amino acids are assembled into proteins using the DNA "recipe".

 ACC CTG CTG GAC CTG GGG GCT TCA CCT GAC TAC AAG GAC AGC CGC
 Thr Leu Leu Asp Leu Gly Ala Ser Pro Asp Try Lys Asp Ser Arg

The building of the protein is complete when the DNA sequence reaches a STOP codon. There are actually three different stop codons: TGA, TAA, and TAG. Any one of them will end the assembly of the protein. There is also a start codon indicating where the gene begins: ATG.

A genetic report may have a section that starts with "c.". The "c." locates an unexpected change in the DNA code (nucleotides). The report may also have a part that starts with "p.", which locates the change in the protein (amino acids) caused by the change in DNA.

For example: variant: c.2594G>A

The nucleotide G has been changed to the nucleotide A at location 2594. This substitution can have 3 possible effects depending upon the sequence at 2594.

Original sequence	Original amino acid	New sequence	New amino acid	Impact
ACG	Thr	ACA	Thr	No effect
ATG	Met	ATA	Ile	Wrong amino acid
TGG	Trp	TGA	STOP	STOP the protein here

The impacts of these three possibilities are likely different. In the first case, there is usually no impact. In the second case, the change in amino acid may be unimportant or it may play a critical role. In the last case, the STOP code will truncate the protein, which usually makes the protein unusable. All three cases are called SNPs[1].

Variant: c.2594_2597dup
Variant: c.2594_2595del

The underscore bar here means that the change affects the range of these positions in the code. "Dup" means the letters at the positions between the two numbers have been duplicated. "Del" means they have been deleted.

---

[1] SNP stands for single-nucleotide polymorphism

Any time there is a change that isn't three letters long (or a multiple of three) there is another important problem: the change shifts the grouping of letters in the reading of the code, which garbles the code. This is called a frameshift. Frameshifts usually cause major disruptions of the gene.

In the example of a stretch of *SHANK3* above, a deletion of two nucleotides looks like this, if the two missing letters are from the 2nd group of three (highlighted in red):

Original Sequence
ACC CTG CTG GAC CTG GGG GCT TCA CCT GAC TAC AAG GAC AGC CGC
Thr Leu Leu Asp Leu Gly Ala Ser Pro Asp Try Lys Asp Ser Arg

Changed Sequence
(CT are dropped, and so the grouping of all letters is automatically shifted).
The bottom row shows the new amino acid sequence:

ACC GCT GGA CCT GGG GGC TTC ACC TGA   CTA CAA GGA CAG CCG
Thr   Ala   Gly   Pro   Gly   Gly   Phe   Thr   STOP   ---- the rest is ignored ----

When the protein is made following this changed code, it has two problems: it has the wrong amino acids, and the TGA will cause a "STOP" to signal the termination of the protein before it should have ended.

p.Leu864Ala
This tells us the Leu (Leucine) amino acid of the original sequence was changed to an Ala (Alanine) amino acid. In some cases, this would be the only change. In a frameshift, it marks the first of a string of changes.

fs*22
FS means frameshift. A frameshift is when the groups of 3 nucleotides have been shifted by gaining or losing 1, 2, 4 or any number not divisible by 3, as in our example. The *22 shows how many more nucleotides are transcribed until it runs into a STOP code. Once it hits the STOP, the rest of the gene is ignored.

Heterozygous
The two copies of *SHANK3* are different. This can result because the copy of *SHANK3* inherited from the mother is different from the copy inherited from the father, or because one copy is typical and the other is an unusual variant.

Homozygous
Both copies of *SHANK3* are the same. Usually that means both copies are typical, because humans cannot survive if both copies are pathogenic.

Part 3: How to understand the technical language for a deletion

In PMS most - but not all - chromosomal deletions remove part or all of *SHANK3*. *SHANK3* is thought to be very important, but it is only one gene. A PMS chromosomal deletion is often large enough to remove 45 to 50 genes. Twelve of these genes have been identified as likely important to the PMS phenotype. Genetic reports for chromosomal deletions use different terms than the ones described above for variants.

Reference genome
Once the human DNA was decoded ("sequenced") for many different people, scientists assembled a master genome (table of the normal sequence of our DNA code), called a reference genome. Over the years the reference has been refined. Many current reports are based on GRCh37/hg19. (Hg19 stands for human genome reference #19). The most recent "build" of the human genome is GRCh38/hg38.

Location of copy number variant
The location of the copy number variation can be described in different ways.

Example 1: terminal deletion
Arr[GRCh37] 22q13.31q13.33(44802524-51304566)x1

Arr
Indicates that the results were obtained by array testing.

22q13.31q13.33
This should look a little familiar. 22q13 means chromosome 22, long arm "q", region 1, band 3. Bands are alternating light and dark regions along the chromosome that are distinguishable from adjacent segments. Bands have sub-bands. In this case, the report says the arrangement affects 22q13.31 to 22q13.33. (sub bands 31 to 33).

(44802524-51304566)
This is the precise location on chromosome 22. The numbers make sense only for a specific reference genome. In this case, the numbers come from hg19.

x1
This says the chromosome array test found the sequence for only one chromosome in that region. It is a good indication that a deletion has occurred in that region on the other chromosome.

Deletion size
Because the element (44802524-51304566) provides exact locations, we can subtract the two numbers to calculate the deletion size. In this example, the deletion is of 6,502,042 base pairs (6.5 Mb). (Since nucleotides are in pairs, the count of base pairs is the same as the count of nucleotides.)

Terminal deletion

When the deleted part of the chromosome extends to the very end of the chromosome (the "terminus"), the deletion is called a terminal deletion.

Interstitial deletion
When the deletion does not extend to the end of the chromosome.

OMIM genes
Genes that have been implicated in various disorders have been cataloged on-line. OMIM is the "Online Mendelian Inheritance in Man" web site. It is a great resource for reading about genes and disorders associated with genes. https://www.omim.org/

Inheritance
When looking up genes it's important to pay attention to the "inheritance". AR stands for autosomal recessive, which generally means that, even if a person has one abnormal variant, having one "normal" (common) variant is sufficient to prevent the disorder. AD stands for autosomal dominant, which generally means that having one abnormal variant is enough to cause the disorder, even if the person also has a normal variant. Phelan McDermid syndrome is an autosomal dominant disorder.

There are other severe disorders related to genes in the same region of chromosome 22, which can occur in people with PMS only if, after one copy of a gene is deleted, the remaining copy is a pathogenic variant. It is rare for people with PMS to also have a pathogenic variant on the second copy of a gene.

46,XY
This is a normal karyotype for a male obtained by chromosome analysis.

Example 2: ring chromosome
46,XY,r(22)(p13q13)

r(22)(p13q13)
r(22) means chromosome 22 has formed a ring, which is abnormal. A ring chromosome forms from deletion of the two terminal ends of the chromosome followed by fusion of the broken ends. On chromosome 22, p13 is on the end of the short arm and q13 is on the end of the long arm. It is common that a ring is missing part of the chromosome on one or both ends. If the q13 end is missing part of the chromosome, that would be a 22q13.3 deletion.

Example 3: translocation
46,XX,t(19;22)(q13.42;q13.31)

t(19;22)(q13.42;q13.31)
This is a translocation between the long arm (q) of chromosome 19 and the long arm (q) of chromosome 22. It is a balanced translocation. No deletion (loss of genetic material) is indicated. A parent with this arrangement can be a "carrier" of PMS. The offspring may (~25% chance) inherit an unbalanced translocation. (see the next example)

der(22)t(22;19)(q13.31;q13.42)

This unbalanced translocation creates a derivative (abnormal) chromosome 22 made from most of chromosome 22 and some of chromosome 19. This represents loss of part of chromosome 22 (chromosome 22 deletion), plus an extra copy of part of chromosome 19 (trisomy). It is a form of PMS.

Part 4: Additional resources

Cytogenetics: Nomenclature and Disease
https://www.cibmtr.org/Meetings/Materials/CRPDMC/Documents/2008/Nov2008/NavarroW_Cytogenetic.pdf

Here is a web site with much more detail.
https://www.hgvs.org/mutnomen/standards.html

The human genome browser locates genes and many other features of each chromosome. Here is the browser set to view a 1.1 Mb terminal deletion of chromosome 22.
https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr22%3A50200719%2D51304566&hgsid=1050460065_ST1ARvaaMs1xzwiLxL6GNuIlef2k

Acknowledgement: this guide would not have been possible without the expertise and patient help of Andy Mitz, and the generosity of Katy Phelan, who took the time to review and comment on an earlier draft. It also is informed by the hard work and creative care of the dedicated researchers and clinicians who focus on the challenges facing our families, and by the advocacy, education and innovation driven by the PMS Foundation.