Notes on ClinGen DM Documentation

ReferenceSequence

- Looks good

ReferenceGenome

- Looks good

GenomicReferenceSequence

- Looks good

TranscriptReferenceSequence

- Look good

ProteinReferenceSequence

- Looks good

GenomicSimpleAllele

- Looks good

TranscriptSimpleAllele

1. Now that we have agreed to remove transcript_variant (type) from version 1.0, I just want to clarify that we allow annotation of the feature effected, not the variation observed (at the transcript level), example:
    a. intron (SO:0000188) —is_a—> primary_transcript_region —is_a—> … —is_a—> sequence_feature (SO:0000110)
    b. five_prime_UTR —is_a—> … —is_a—> sequence_feature (SO:0000110)
    c. And not:
        i. intron_variant (SO:0001627) —is_a—> transcript_variant (SO:0001576).
        ii. 5_prime_UTR_variant —is_a—> transcript_variant (SO:0001576).
    d. Maybe a change in the wording would be (from/to):
        i. *"While TranscriptSimpleAllele is defined with respect to a TranscriptReferenceSequence, the allele does not necessarily lie within the TranscriptReferenceSequence itself. For instance, intronic variants can be defined with respect to a transcript sequence, even though they are not part of the sequence. It would also be appropriate to represent a variant upstream or downstream of a transcript using a TranscriptSimpleAllele in this fashion"*
        ii. While TranscriptSimpleAllele is defined with respect to a TranscriptReferenceSequence, the allele does not necessarily lie within the TranscriptReferenceSequence itself. For instance, intronic variants can be described with respect to a transcript feature in which the variant lies, even though they are not part of the sequence. In version1.0, describing a resulting molecular consequence from a transcript variant would be handed at the ProteinSimpleAllele level.

GenomicSimpleAllele

- Looks good

ProteinSimpleAllele

- Looks good
    - This page was not added, so I did. I used the text from GenomeSimpleAllele and made text changes to protein; please amend.

NucleotideSimpleAllele

- No changes needed

SimpleAllele

- This is all great and I agree overall with with how it's worded but I have one thought:
    - Given that we state "*SimpleAllele must be either a GenomicSimpleAllele, a TranscriptSimpleAllele, or a ProteinSimpleAllele*", would each sub-class benifit from having a "type(Effect/MC)" and "feature" field.

AlleleName & CanonicalAlleleIdentifier

- Looks good.

1. **TODO**: misspelled on the VP model.

CanonicalAllele

- Just for my clarification when we state this: *"Furthermore, if a genomic variant occurs in a transcript, the allele may be defined with respect to either the genomic sequence, or the transcript sequence."* we are assuming annotations at two separate sub-classes, TranscriptSimpleAllele and GenomicSimpleAllele, correct.