An Open-Access\* Catalog of Argumentation Schemes in

Biomedical Genetics Research Articles

February 26, 2018

Nancy L. Green

Department of Computer Science

University of North Carolina Greensboro

Greensboro, NC 27502

nlgreen@uncg.edu

Table of Contents

1. Introduction
2. Definitions of domain terminology
3. Ontology of argument schemes:
   1. Causation
      1. One-group
         1. Agreement arguments
            1. Method of Agreement
            2. Failed Method of Agreement (no Effect)
            3. Failed Method of Agreement (no Cause)
         2. Eliminate Candidates
         3. Explanation-based
            1. Effect to Cause
            2. No Effect to No Cause
            3. Consistent with Predicted Effect
      2. Two-group
         1. Difference arguments
            1. Method of Difference
            2. Eliminate Difference
            3. Failed Method of Difference
         2. Analogy (Causal)
         3. Explanation-based
            1. Consistent Explanation
            2. Difference Consistent Explanation
   2. Other
      1. Classification
      2. Confirmation
4. Acknowledgements
5. References

\*This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

1. **Introduction**

This catalog describes some types of arguments, or argument/argumentation schemes (Walton et al. 2008), that we have identified by human analysis of articles from biomedical journals. The articles are on the topic of genetic variants (mutations) that may have an adverse effect on human health.

An argument scheme consists of a generalized description of the *premises* (reasons) supporting a *conclusion*. Argument schemes can be described in a non-domain-specific manner such as the following scheme related to Mill’s Method of Agreement, a form of scientific reasoning noted by Mill in the nineteenth century (Jenicek and Hitchcock 2005).

**Scheme:** Method of Agreement

**Premise:** A group of individuals I have atypical property P

**Premise:** All of the individuals in I have atypical feature F

**Conclusion:** F may be the cause of P (in I).

As in the above example, the conclusions of argument schemes in this document are not necessarily deductively valid nor asserted with complete certainty. Also, note that in some cases a premise or conclusion may not be stated explicitly, and so is marked ‘implicit’ in this catalog.

Since the goal of this version of the catalog was to make it easier for human annotators to identify arguments in this particular domain, the argument schemes are described here using domain terminology. (The 8/24/2015 version of this catalog (Green 2015) described them in non-domain-specific terms. However, we expect that the schemes in this catalog can be used in other domains of interest in biology by a similar process of adaptation.) For example, the above scheme is described in this catalog as follows:

**Scheme:** Method of Agreement

**Premise:** A group of individuals I have atypical phenotype P

**Premise:** All of the individuals in I have atypical genotype M.

**Conclusion:** M may be the cause of P (in I).

Domain terms such as *phenotype* and *genotype* are defined in the next section.

For further information on the schemes and their potential use in argument mining, see (Green to appear in 2018). For information on two studies that evaluated some schemes in earlier versions of the catalog, see (Green 2015; Green 2017).

Each scheme definition is accompanied by one or more examples. The sources of examples are given in the References section. Note that articles from the CRAFT[[1]](#footnote-1) corpus are open-access [Bada et al. 2012; Verspoor et al. 2012]. We have annotated arguments in some articles in the CRAFT corpus. Those annotated articles will be made available soon as a companion to this document.

1. **Definitions of domain terminology**

**Genotype** describes a variation at the level of chromosome, region on the chromosome, or gene that may have a deleterious effect on organisms with that genotype. Sometimes, instead of one genotype, the article may refer to a set of candidate genotypes.

**Gene Product** describes a gene product, such as mRNA or protein, in terms of expression, level, or form, or in terms of function or pathway.

**Group** describes one or more individuals of the same taxon (species) who have a property in common, e.g., the affected members of a family, the unaffected members of a family, a control group, individuals who have a certain mutation, etc. (Note that for the sake of simplicity in the definitions, a group may have only one member.)

**Phenotype** describes deleterious effects of a genotype on an organism’s body; it includes various types of effects such as hearing loss, developmental delay, movement disorder, abnormal levels of a certain substance in the blood, and inability of an organ to perform its function. Sometimes the effects are classified as a disease (SCA15) or a group of related diseases (ataxia).

**Segregate/Segregating** refers to when inheritance of a phenotype in a family is consistent with autosomal dominant or recessive (Mendelian) inheritance.

1. **Ontology of argument schemes**

The schemes have been organized hierarchically to highlight relationships among the schemes. Here is a brief description of their organization and their relation to more general descriptions in earlier argumentation studies.

Most of the schemes are under (3.1 Causation) since the conclusions of these schemes are assertions involving causation. The schemes in (3.1) are then divided into (3.1.1 One-group) and (3.1.2 Two-group), depending upon whether one group of individuals is studied, or two groups of individuals are compared, respectively.

Under (3.1.1 One-group), the (3.1.1.1 Agreement) arguments are similar in some ways to Mill’s Method of Agreement, described in section 1. The category (3.1.1.2 Eliminate Candidates) seems to be a specialization of the disjunctive syllogism. The category (3.1.1.3 Explanation-based) contains schemes whose premises include some sort of causal explanation. Effect to Cause (3.1.1.3.1) is related to the more general Effect to Cause scheme (e.g. Walton et al. 2008).

The Two-group classification (3.1.2) is divided into Difference arguments (3.1.2.1), Analogy (Causal) (3.1.2.2), and Explanation-based schemes (3.1.2.3). The Difference arguments are similar in some ways to Mill’s Method of Difference, another form of scientific reasoning noted by Mill in the nineteenth century (Jenicek and Hitchcock 2005). Analogy (Causal) (3.1.2.2) is a specialization of Analogy (e.g. Walton et al. 2008), in which the conclusion is a causal assertion. The premises of Explanation-based schemes (3.1.2.3) include some sort of causal explanation that applies to the two groups under study.

Finally, the Other (3.2) category includes arguments not about causation but about how a certain classification is defined (3.2.1 Classification), or the reliability of data (3.2.2 Confirmation). Classification (3.2.1) is a specialization of an argument from classification (Walton et al. 2008).

* 1. **Causation**
     1. **One-group**
        1. **Agreement arguments**

**Argumentation Scheme: Method of Agreement**

**Premise:** A group of individuals I have atypical phenotype P

**Premise:** All\* of the individuals in I have atypical genotype M, or set of atypical genotypes M1 … Mn.

**Conclusion:** M (or one of M1 … Mn) may be the cause of P (in I)

\*Note: In some cases, an article makes a Method of Agreement argument even though a small fraction of the individuals do not have the genotype(s) in question. In these cases the author apparently considered the number of exceptions insignificant and/or the phenotype of the exceptional individuals could be due to other factors.

Example 1 (CRAFT 175900787):

We obtained genomic DNA from three affected family members ... from the kindred originally used to define and map SCA15 ... We performed two experiments concurrently in three affected members of this family: sequence analysis of the coding exons of ITPR1 and high-density genome-wide SNP genotyping ... visualization of log R ratio and B allele frequency metrics from the genome-wide SNP genotyping experiments clearly showed data consistent with a heterozygous genomic deletion across the first one-third of ITPR1 and across the first half of a neighboring gene, SUMF1 ...

**Paraphrase of argument:**

**Argumentation scheme: *Method of Agreement***

**Premise:** A group (three family members) was affected with SCA15.

**Premise:** The group had a heterozygous genomic deletion in ITPR1-SUMF1

**Conclusion (implicit):** This deletion may be the cause of SCA15 in this group.

Example 2 (Jameel et al. 2014):

Whole exome enrichment was performed on DNA samples from the two affected family members... Using ... data from the two brothers, we identified altogether nine novel candidate variants ...

**Paraphrase of argument:**

**Argumentation scheme: *Method of Agreement***

**Premise:** Two family members were affected [with cerebral palsy (CP) symptoms]

**Premise:** The two affected family members had nine candidate variants

**Conclusion (implicit):** One of the nine novel variants may be the cause [of the CP symptoms]

**Argumentation Scheme: Failed Method of Agreement (no Effect)**

(called *Argument by Negative Association* in 8/24/2015 Catalog)

**Premise:** No one in a set of individuals I has atypical phenotype P

**Premise:** The individuals in I have atypical genotype G

**Conclusion:** G is not likely a/the cause of P

Example (CRAFT 175900787):

It is improbable that heterozygosity for the deletion of SUMF1 ... of itself causes or contributes to SCA15. Homozygous mutation of SUMF1 results in autosomal recessive multiple sulfatase deficiency ... No co-occurrence of ataxia has been described in (heterozygous) parents of patients with multiple sulfatase deficiency.

**Paraphrase of argument:**

**Argumentation scheme: *Failed Method of Agreement (no Effect)***

**Premise (partly implicit):** Parents of patients with autosomal recessive multiple sulfatase deficiency are heterozygous for deletion of SUMF1. (According to Mendelian inheritance principles, the children must have inherited one SUMF1 mutation from each parent, and since the parents don’t have multiple sulfatase deficiency themselves, they must be heterozygous for deletion of SUMF1.).

**Premise:** No co-occurrence of ataxia has been described in heterozygous parents of patients with multiple sulfatase deficiency.

**Conclusion:** It is improbable that heterozygosity for the deletion of SUMF1 is a/the cause of SCA15.

**Argumentation Scheme: Failed Method of Agreement (no Cause)**

(Called *Failed Method of Agreement*in 8/24/2015 Catalog)

**Premise:** A set of individuals I have atypical phenotype P

**Premise:** Not all (or none) of the individuals in I have atypical genotype M

**Conclusion:** M is not likely a/the cause of P

Example (Schrauwen et al. 2012):

We screened 24 unrelated affected Belgian and Dutch individuals with a moderate to severe hearing loss for mutations in *CABP2* ..., but we could not identify a clear damaging mutation in any of them.

**Paraphrase of argument:**

**Argumentation scheme: *Failed Method of Agreement (no Cause)***

**Premise:** 24 unrelated Belgian and Dutch individuals had moderate to severe hearing loss.

**Premise:** They did not have a CABP2 mutation.

**Conclusion (implicit):**  A CABP2 mutation is not likely the cause of their hearing loss.

**3.1.1.2. Eliminate Candidates**

**Argumentation Scheme: Eliminate Candidates**

(revised from *Eliminate Causal Candidates* in 8/24/2015 Catalog)

**Premise:** A set CC of candidate genotypes has been identified (in a group or an individual) one of which may be the cause of atypical phenotype P (in that group or individual).

**Premise:**  One or more genotypes in CC can be eliminated as candidates because (a) the genotype is a frequent polymorphism, or (b) is not predicted as pathogenic (e.g. non-synonymous variants), or (c) is not previously associated with P, or (d) the other genotypes in CC are better candidates because they are more serious types of mutation (e.g. missense or nonsense), or (e) there is other evidence for one of the candidates so the other previous candidates can be eliminated.

**Conclusion:** One of the remaining members of CC may be the cause of P (in that group or individual).

Example (Jameel et al. 2014):

... the enriched DNA was sequenced ... Common variants were excluded by filtering against dbSNP130 … and 800 inhouse exome. Using filtered WES data from the two brothers, we identified altogether nine novel candidate variants. Five missense variants were located in the X-chromosome genes … However, none of the amino acid substitutions were predicted as pathogenic and the five genes were not previously associated with the clinical features observed in our patients. Four autosomal and homozygous variants were identified ...

**Paraphrase of 2 arguments in above excerpt:**

**Argumentation scheme: *Eliminate Candidates***

**Premise (implicit):** A set of candidate genotypes CC responsible for the patients’ phenotype has been identified by DNA sequencing.

**Premise:** Common variants (polymorphisms) that were found in dbSNP130 and 800 inhouse exomes were eliminated from CC.

**Premise**: The remaining nine genotypes in CC may be the cause of the patients’ phenotype.

**Argumentation scheme: *Eliminate Candidates***

**Premise:**  The remaining nine genotypes in CC may be the cause of the patients’ phenotype.

**Premise:** Five variants in CC were eliminated because they were not predicted to be pathogenic and were not previously associated with the patients’ phenotype.

**Conclusion (implicit):** The remaining four autosomal and homozygous variants may be the cause of the patients’ phenotype.

*Eliminate Candidates* examples continued on next page

Example (Jameel et al. 2014):

Whole exome enrichment was performed on DNA samples from the two affected family members [with cerebral palsy] ... Four autosomal and homozygous variants were identified in the TNFRSF14, AP4M1, RGMA and NINL genes.

**Paraphrase of argument**:

**Argumentation scheme:** *Eliminate Candidates*

**Premise** (from Background section): Mutations in either AP4B1, AP4E1, AP4M1 or AP4S1 may cause cerebral palsy.

**Premise:** Two brothers with cerebral palsy have candidate mutations in the TNFRSF14, AP4M1, RGMA and NINL genes.

**Conclusion** (implicit): The AP4M1 may be the cause of cerebral palsy in the two brothers.

* + - 1. **Explanation-based**
         1. **Effect to cause**

**Argumentation Scheme: Effect to Cause**

(combines *Effect to Cause* and *Effect to Causal* *Candidates* of 8/24/15 catalog)

**Premise:** An individual or group of individuals I have atypical phenotype P

**Premise:** Atypical genotype M (or M in a set of candidate genotypes M1…Mn) has a possible causal link to P.

**Premise:** It is not known if I has M.

**Conclusion**: (I may have M and) M may be the cause of P in I.

Example (Schrauwen et al. 2012):

The DFNB93 region contains ... several genes [which] are expressed in the mouse and human inner ear. Because there are many strong candidates in the region, we sequenced all genes and noncoding genes in this region ... to identify the disease-causing mutation in one affected individual [V:14] from the family

**Paraphrase of argument**:

**Argumentation scheme:** *Effect to Cause*

**Premise** (from previous text):A certain individual (V:14) has hearing loss.

**Premise**: Several genes in the DFNB93 region are expressed in the ear.

**Premise** (domain knowledge): A variant of a gene that is expressed in a tissue or system may lead to an abnormality in that tissue/system.

**Conclusion**: Individual V:14 may have a variant in the DFNB93 region and it may be the cause of V:14’s hearing loss.

Example 2 (Baumann et al. 2012):

EDS VIA “… is characterized by severe muscle hypotonia at birth, progressive kyphoscoliosis, … and joint hypermobility.” [several paragraphs later] “The index person P1 … was referred … for the evaluation of severe kyphoscollosis, joint hypermobility, and muscle weakness. He was initially suspected to have EDS VIA …”

**Paraphrase of argument**:

**Argumentation scheme:** *Effect to Cause*

**Premise**:P1 has severe kyphoscollosis, joint hypermobility, and muscle weakness.

**Premise**: EDS VIA is characterized by severe muscle hypotonia at birth, progressive kyphoscoliosis, … and joint hypermobility.

**Conclusion:** P1 may have EDS VIA.

**3.1.1.3.2 No Effect to No Cause**

**Argumentation Scheme: No Effect to No Cause**

**Premise:** Atypical genotype M typically leads to atypical phenotype P.

**Premise:** An individual I does not have P

**Conclusion**: It is not likely that I has M.

Example 2 (Baumann et al. 2012):

“… an increased LP/HP ratio, which is diagnostic for EDS VIA.” [several paragraphs later: “[He] was initially suspected to have EDS VIA, but the urinary LP/HP ratio was in the normal range.”

**Paraphrase of argument**:

**Argumentation scheme:** *No Effect to No Cause*

**Premise:** An individual with the mutation for EDS VIA usually has an increased LP/HP ratio.

**Premise:** This individual’s LP/HP ratio was normal.

**Conclusion:** This individual does not have the mutation for EDS VIA.

**3.1.1.3.3 Consistent with predicted effect**

Note that this explanation-based scheme is similar to *Consistent Explanation* and *Difference Consistence Explanation*: However, those argument schemes explicitly compare two groups, but this scheme is applied to only one group.

**Argumentation Scheme: Consistent with Predicted Effect**

**Premise:** Some group G has atypical genotype M and atypical phenotype P.

**Premise:**  There is a predicted causal mechanism at the molecular level that explains how M could result in P.

**Conclusion:** M may be the cause of P in G.

Example (Schrauwen et al. 2012):

On the basis of our present findings ... dysregulation of IHC synaptic transmission could be one pathogenic mechanism underlying hearing impairment in DFNB93 ... In IHCs, the c.637+1G>T mutation in *CABP2* would most likely enhance inactivation of synaptic Ca2+ influx. This, in turn, could reduce rates of transmitter release and consequently diminish spiral ganglion neuron firing and ascending auditory-pathway activation.

**Paraphrase of argument**:

**Argumentation scheme:** *Consistent with Predicted Effect*

**Premise** (from preceding text):Certain individuals with hearing loss have a c.637+1G>T mutation in *CABP2.*

**Premise:** The c.637+1G>T mutation in *CABP2* would most likely enhance inactivation of synaptic Ca2+ influx. This, in turn, could reduce rates of transmitter release and consequently diminish spiral ganglion neuron firing and ascending auditory-pathway activation.

**Premise** (from preceding text): the auditory system ... involves coordination of multiple processes involving different parts of the ear and nervous system. A defect in any part of this complex chain of events can lead to hearing impairment or deafness.

**Conclusion:** The c.637+1G>T mutation in *CABP2* may be the cause of hearing loss in those individuals.

* + 1. **Two-group**

**3.1.2.1 Difference Arguments**

**Argumentation Scheme: Method of Difference**

**Premise:** A group of individuals I have atypical phenotype P

**Premise:** All\* of the individuals in I have atypical genotype M

**Premise:** A group of individuals C do not have P.

**Premise:** None\* of the individuals in C have M.

**Conclusion:** M may be the cause of P (in I)

\*N**ote:** An article could make a Method of Difference argument where “all” is actually “approximately all”, and/or “none” is actually “approximately none”. In these cases the author apparently considers the number of exceptions insignificant.

Example (CRAFT 175900787):

With three cerebellar ataxia families segregating\*\* a SUMF1–ITPR1 deletion, and this deletion not observed in a control population, we may reasonably conclude that the association is causal, and that the deletion is indeed the genetic basis of the disease ...

**Paraphrase of 1st argument**:

**Argumentation scheme:** *Method of Difference*

**Premise:** In three families, the family members with cerebellar ataxia had aSUMF1–ITPR1 deletion

**Premise:** This deletion was not observed in a control group.

**Conclusion:** A SUMF1-ITPR1 deletion may be the cause of SCA15 in the affected members of the three families.

**Paraphrase of 2nd argument**:

**Argumentation scheme:** *Method of Difference\*\**

**Premise:** In three families, cerebellar ataxia segregated with aSUMF1–ITPR1 deletion (i.e. family members with cerebellar ataxia had the deletion, and those without cerebellar ataxia did not have the deletion).

**Conclusion:** A SUMF1-ITPR1 deletion may be the cause of SCA15 in the three families.

**\*\*Note :** In addition to the second Method of Difference argument paraphrased here, there is also a domain-specific argument that inheritance of the disorder was consistent with Mendelian inheritance of the deletion.

**Argumentation Scheme: Eliminate Difference**

**Premise:** Group G1 has atypical phenotype P

**Premise:** Group G1 has atypical genotype M1M2

**Premise:** Group G2 does not have phenotype P

**Premise:** Group G2 has atypical genotype M2

**Premise:** M2 does not cause P.

**Premise:** M1 is the difference between M1M2 and M2.

**Conclusion:** M1 may be the cause of P

Example (CRAFT 175900787):

With three cerebellar ataxia families segregating a SUMF1–ITPR1 deletion... we may reasonably conclude that the association is causal ... It is improbable that heterozygosity for the deletion of SUMF1 ... of itself causes or contributes to SCA15... No co-occurrence of ataxia has been described in (heterozygous) parents of patients with multiple sulfatase deficiency.

**Paraphrase of argument**:

**Argumentation scheme:** *Eliminate Difference*

**Premise:** Group G1 consists of affected members of three cerebellar ataxia families.  
**Premise:** All individuals in G1 have deletion in ITPR1-SUMF1

**Premise:** Group G2 consists of parents of patients with multiple sulfatase deficiency.

**Premise:** Individuals in G2 do not have ataxia.

**Premise:** Individuals in G2 have the SUMF1 deletion (but not ITPR1 deletion).

**Premise**: SUMF1 deletion does not cause ataxia.

**Premise** (domain knowledge)**:** ITPR1 is the difference between SUMF1 and ITPR1-SUMF1.

**Conclusion** (implicit): ITPR1 deletion may be the cause of ataxia in G1.

**Argumentation Scheme: Failed Method of Difference**

**Premise:** A group of individuals I have atypical phenotype P

**Premise:** All\* of the individuals in I have atypical genotype M

**Premise:** A set of individuals C do not have P.

**Premise:** At least one of the individuals in C have M.

**Conclusion:** M is not likely the cause of P (in I)

\*N**ote:** An article could make a Failed Method of Difference argument where “all” is actually “approximately all”. In these cases the author apparently considers the number of exceptions insignificant.

Example (Charlesworth et al. 2012):

This strategy revealed three potentially pathogenic variants. The first, a heterozygous frameshift deletion ... in exon 2 of *TBC1D7 ...*, failed to fully segregate, given that individual II-5, who is unaffected at age 61, and individual III-8, who is unaffected at age 32, exhibit the deletion.

**Paraphrase of argument**:

**Argumentation scheme:*****Failed Method of Difference***

**Premise:** All of the affected individuals had the mutation.

**Premise:** Two unaffected individuals have the mutation.

**Conclusion (implicit):** The mutation is not likely the cause of the condition.

**3.1.2.2 Analogy (Causal)**

Note how this differs from Consistent with Predicted Effect and Consistent Explanation: no potential causal mechanism is described in the premises of the following scheme.

**Argumentation Scheme: Argument by Analogy**

Premise: Phenotype P1 of group G1 is similar to phenotype P2 of group G2

Premise: In group G1, genotype M1 may be the cause of P1.

Premise: Group G2 has genotype M2 that is similar to M1.

Conclusion: M2 may be the cause of P2.

Example (CRAFT175900787):

Given our interest in human neurological disease we sought to identify any cognate human disorders where linkage had been established to the syntenic region of the human genome, but where no causal mutation had been identified. SCA15, an adult-onset autosomal dominant progressive ataxia is linked to this locus [5]. Although missense mutation of ITPR1 had previously been ruled out [2] and the mode of inheritance was inconsistent with that seen in the Itpr1Δ18 and Itpr1opt mice, the phenotypic presence of ataxia in the mice led us to reexamine this candidate gene as a possible cause of SCA15.

**Paraphrase of argument**:

**Argumentation scheme: *Argument by Analogy\****

**Premise:** Two groups of mice with Itpr1 mutations (Itpr1Δ18 and Itpr1opt mice) had an ataxia phenotype.

**Premise:** The SCA15 phenotype in humans is similar to the ataxia phenotype in mice.

**Premise (implicit; conclusion of a previous argument)**: The Itpr1 mutation may be the cause of the ataxia phenotype in those mice.

**Premise:** Itpr1 in mice is similar (i.e. syntenically related) to ITPR1 in humans.

**Conclusion**: An ITPR1 genetic mutation may be the cause of SCA15 in humans.

\*Note: The analogy can be based on different taxa as in this example (humans and mice) or based on different groups of the same taxon (such as two groups of mice with different genotypes).

**3.1.2.3 Explanation-based**

Note how this differs from Analogy: no potential causal mechanism is described in the premises of Analogy, while the premises of Consistent Explanation describe such a mechanism. Also, this differs from Consistent with Predicted Effect in that this scheme explicitly compares two groups.

**Argumentation Scheme: Consistent Explanation**

**Premise:** Group G1 has atypical genotype M1, atypical phenotype P1, atypical gene product Prod

**Premise:** M1 produces Prod, which is associated with P1, or results in defective function F that could lead to P1

**Premise:** M1 may cause P1

**Premise:** Group G2 has atypical genotype M2, atypical phenotype P2, atypical gene product Prod

**Premise:** M1 is similar to M2 and P1 is similar to P2

**Premise:** M2 produces Prod, which is associated with P2, or results in defective function F that could lead to P2

**Conclusion:** M2 may cause P2 in G2.

Example (CRAFT 175900787): As with the Itpr1opt/opt mice ... the in-frame Itpr1Δ18/Δ18 deletion mutation results in markedly decreased levels of Itpr1 in cerebellar Purkinje cells. In these two spontaneous mutants [1] ... decreased Itpr1 expression is associated with the same autosomal recessive movement disorder ...

**Paraphrase of argument**:

**Argumentation scheme:** *Consistent Explanation*

**Premise:** Group G1 of mice has atypical Itpr1opt/opt genotype, an autosomal recessive movement disorder, and decreased Itpr1 protein expression in G1’s cerebellar Purkinje cells.

**Premise:** Itpr1opt/opt mutation produces decreased Itpr1 protein expression, which is associated with the autosomal recessive movement disorder in G1.

**Premise** (from previous argument): Itpr1opt/opt mutation may cause the autosomal recessive movement disorder in G1.

**Premise:** Group G2 of mice has atypical Itpr1Δ18/Δ18 genotype, the same autosomal recessive movement disorder as in G1, and decreased Itpr1 protein expression in G2’s cerebellar Purkinje cells.

**Premise:** Itpr1Δ18/Δ18 mutation produces decreased Itpr1 protein expression in cerebellar Purkinje cells in G2, which is associated with the autosomal recessive movement disorder in G2

**Premise:** Itpr1opt/opt mutation is similar to Itpr1Δ18/Δ18 mutation

**Conclusion**: The Itpr1Δ18/Δ18 mutation may cause the autosomal recessive movement disorder in G2.

**Argumentation Scheme: Difference Consistent Explanation**

**Premise:** Group G1 has atypical genotype M, atypical phenotype P, atypical gene product Prod

**Premise:**  Group G2\* does not have M, P, Prod

**Premise:** M produces Prod

**Premise:** Prod is associated with P, or results in defective function F that could lead to P

**Conclusion:** M may cause P in G1.

\*Note: G1 and G2 can be from different taxa such as mouse or human. Also, our domain expert noted that (1) typically evidence from human in vitro experiments is stronger than evidence from mouse in vivo experiments because of the difference between taxa, and evidence from human in vivo experiments is stronger than evidence from human in vitro experiments; and (2) between different species protein function is more conserved than protein structure, which is more conserved than protein sequence, which is more conserved than mRNA, which is more conserved than gene sequence.

Example (CRAFT 175900787):

Conversely, mutation of ITPR1 is biologically plausible as a cause of [human] ataxia: the protein is highly expressed in Purkinje cells in mice; as we have shown here, mice with mutation at this locus present with ataxia; and perturbed Ca2+ signaling has previously been implicated in the etiology of ataxia, notably in episodic ataxia type 2 and SCA6 [8]. In further support of this conclusion, analysis of[human]protein levels of ITPR1 in Epstein-Barr virus (EBV) immortalized lymphocytes from affected and unaffected AUS1 family members revealed that all affected members showed a dramatic decrease in ITPR1 levels

**Paraphrase of argument**:

**Argumentation scheme:** *Difference Consistent Explanation*

**Premise:** Group G1 of affected AUS1 family members have ITPR1 mutation, ataxia phenotype, and decreased expression of ITPR1 protein in Purkinje cells

**Premise:** Group G2 of unaffected AUS1 family members do not have ITPR1 mutation, ataxia phenotype, or decreased expression of ITPR1 protein in Purkinje cells

**Premise (in preceding text)**: Itpr1Δ18/Δ18 and Itpr1opt/opt mutation produces decreased Itpr1 protein level in cerebellar Purkinje cells

**Premise (domain knowledge):** Decreased Itpr1 protein level in Purkinje cells results in perturbed Ca2+ signaling

**Premise:** Perturbed Ca2+ signaling is associated with ataxia

**Conclusion:** The ITPR1 mutation may cause ataxia in affected AUS1 family members.

* 1. **Other**

**3.2.1 Classification**

**Argumentation Scheme: Classification**

**Premise:** A group of individuals have certain characteristics.

**Premise:** Having those characteristics is classified as disease or disorder D.

**Conclusion:** The individuals have D.

Example (CRAFT 16121255):

... we found a family of mice that urinated and drank excessively. Serum and urine analysis showed that plasma glucose levels were normal and there was no glucose in the urine (unpublished data). Hence, this was an example of diabetes insipidus.

**Paraphrase of argument:**

**Argumentation scheme:** *Classification*

**Premise**: A group of mice urinated and drank excessively, but had normal plasma glucose levels, and no glucose in their urine.

**Premise:** Diabetes insipidus is characterized by excessive urination and thirst

**Premise**: Diabetes insipidus is characterized by normal plasma glucose levels, and no glucose in urine.

**Conclusion:** The group of mice had diabetes insipidus.

**3.2.2 Confirmation**

**Argumentation Scheme: Confirmation**

**Premise:** Some data D has been established by one method.

**Premise:** D has been confirmed by another equally or more reliable method, or no data conflicting with D has been established.

**Conclusion:** Confidence in D is higher (than before).

Example (Woo et al. 2014):

Two missense mutations, p.Pro240Leu and p.Glu1595Lys, were identified to be compound heterozygous mutations in the proband of SR-106 family. P.Pro240Leu was confirmed in the patient’s father by Sanger sequencing ... We subsequently sequenced ... however no other mutation was detected.

**Paraphrase of argument**:

**Argumentation scheme:** *Confirmation*

**Premise:** Two missense mutations p.Pro240Leu and p.Glu1595Lys were identified in the proband.

**Premise:** p.Pro240Leu was confirmed in the proband’s father by Sanger sequencing.

**Conclusion:** Confidence in the identification of p.Pro240Leu in the proband is higher (than before).

1. **Acknowledgments**

We thank Michael Branon and Bishwa Giri for their help with the support of a University of North Carolina Greensboro 2016 Summer Faculty Excellence Research Grant.

**References**

Bada, M., Eckert, M., Evans, D., et al. 2012. Concept Annotation in the CRAFT corpus. *BMC Bioinformatics* 13:161.

Baumann et al. 2012. Mutations in FKBP14 Cause a Variant of Ehlers-Danlos Syndrome with Progressive Kyphoscoliosis, Myopathy, and Hearing Loss. *The American Journal of Human* Genetics 90, 201-216, February 10, 2012.

Charlesworth et al. 2012. Mutations in ANO3 Cause Dominant Craniocervical Dystonia: Ion Channel Implicated in Pathogenesis. *The American Journal of Human* Genetics 91, 1041-1050, December 7, 2012.

CRAFT 17590087: van de Leemput, J., et al. Deletion at *ITPR1* Underlies Ataxia in Mice and Spinocerebellar Ataxia 15 in Humans. *PLoS Genetics*, June 2007, Volume 3, Issue 6, e106, 1076-1082.

CRAFT 16121255: Lloyd, D, Halt, FW, Tarantino, LM, and Gekakis, N. Diabetes Insipidus in Mice with a Mutation in Aquaporin-2. *PLoS Genetics*, August 2005, Volume 1, Issue 2, e20, 0171-0178.

Green, N.L. 2015. Identifying Argumentation Schemes in Genetics Research Articles. In *Proc. Second Workshop on Argumentation Mining*. *Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015), May 31-June 5, 2015, Denver, Colorado, USA.*

Green, N.L. 2016. Implementing Argumentation Schemes as Logic Programs. In *Proc. of the Int. Joint Conf. on AI ( IJCAI-16) Workshop on Computational Models of Natural Argument (CMNA-16)*, CEUR-WS 1876.

Green, N.L. 2017. Manual Identification of Arguments with Implicit Conclusions Using Semantic Rules for Argument Mining. In *Proc. Fourth Workshop on Argumentation Mining*. (EMNLP 2017), September 8, 2017, Copenhagen.

Green, N.L. To appear in 2018. Towards Mining Scientific Discourse Using Argumentation Schemes. *Argument and Computation.*

Jameel et al. 2014. A Novel AP4M1 Mutation in Autosomal Recessive Cerebral Palsy

Syndrome and Clinical Expansion of AP-4 Deficiency. *BMC Medical Genetics* 2014,

15:133.

Jenicek, M. and Hitchcock, D. 2005. *Logic and Critical Thinking in Medicine*. American

Medical Association Press, 2005.

Schrauwen et al. 2012. A Mutation in CABP2, Expressed in Cochlear Hair Cells, Causes Autosomal-Recessive Hearing Impairment. *The American Journal of Human Genetics* 91, 636-645, October 5, 2012.

Verspoor, K., Cohen, K.B., Lanfranchi, A., et al. 2012. A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. *BMC Bioinformatics* 2012, 13:207.

Walton, D., Reed, C., and Macagno, F. 2008. *Argumentation Schemes.* Cambridge University Press.

Woo et al. 2014. Identification of CDH23 mutations in Korean families with hearing loss by whole-exome sequencing. *BMC Medical Genetics* 2014 Apr 28;15:46.

1. Available at <http://bionlp-corpora.sourceforge.net/CRAFT/index.shtml> [↑](#footnote-ref-1)