Identifying Argumentation Schemes in Genetics Research Articles

Nancy L. Green
Dept. of Computer Science
U. of N. Carolina Greensboro
Greensboro, NC 27402, USA
nlgreen@uncg.edu

Abstract

This paper presents preliminary work on identification of argumentation schemes, i.e., identifying premises, conclusion and name of argumentation scheme, in arguments for scientific claims in genetics research articles. The goal is to develop annotation guidelines for creating corpora for argumentation mining research. This paper gives the specification of ten semantically distinct argumentation schemes based on analysis of argumentation in several journal articles. In addition, it presents an empirical study on readers' ability to recognize some of the argumentation schemes.

1 Introduction

There has been an explosion in the on-line publication of genetics research articles, creating a critical need for information access tools for genetics researchers, biological database curators, and clinicians. Research on biological/biomedical natural language processing (BioNLP) is an active area of research with the goal of developing such tools. Previous research in BioNLP has focused mainly on fundamental text mining challenges such as named entity recognition and relation extraction (Cohen and Demner-Fushman 2014). However, a key feature of scientific writing is the use of argumentation.

It is important for information access tools to recognize argumentation in scientific text. First, argumentation is a level of discourse analysis that provides critical context for interpretation of a text. For example, a text may give an argument against a hypothesis P, so it would be misleading for a text mining program to extract P as a fact

stated in that text. Second, a user should be able to access a summary of arguments for and against a particular claim. Also, to evaluate the strength of an argument a user should be able to see the arguments upon which it depends, i.e., arguments supporting or attacking its premises. Third, tools that display citation relationships among documents (Teufel 2010) could provide finer-grained information about relationships between arguments in different documents.

Argumentation mining aims to automatically identify arguments in text, the arguments' premises, conclusion and argumentation scheme (or form of argument), and relationships between arguments in a text or set of texts. Most previous work in argumentation mining has focused on non-scientific text (e.g. Mochales and Moens 2011; Feng and Hirst 2011; Cabrio and Villata 2012). Previous NLP research on scientific discourse (e.g. Mizuta et al. 2005; Teufel 2010; Liakata 2012a) has focused on recognizing information status (hypothesis, background knowledge, new knowledge claim, etc.) but has not addressed deeper argumentation analysis.

This paper presents our preliminary work on identification of argumentation schemes in genetics research articles. We define this subtask, unlike some others, e.g. Feng and Hirst (2011), as identifying the premises and conclusion of an argument together with the name of its argumentation scheme. One contribution of the paper is the specification of ten semantically distinct argumentation schemes based on analysis of argumentation in several genetics journal articles. Our goal is to develop annotation guidelines for creating corpora for argumentation mining research. Most of the schemes do not appear in the principal catalogue of argumentation schemes cited in past argumentation mining studies (Walton et al. 2008). In addition, we present an empirical study on readers' ability to recognize some

of the argumentation schemes. The paper concludes with discussion on plans for annotating debate in this type of discourse.

2 Background and Related Work

An argument consists of a set of *premises* and a *conclusion*. *Enthymemes* are arguments with implicit premises and/or conclusions. *Argumentation schemes* are abstract descriptions of acceptable, but not necessarily deductively valid, forms of argument used in everyday conversation, law, and science (Walton et al. 2008). To illustrate, an abductive argumentation scheme is common in medical diagnosis. The premise is that a certain event E has been observed (e.g. coughing). Another, sometimes implicit, premise is that C-type events often lead to E-type events. The tentative conclusion is that a certain event C has occurred (e.g. a respiratory infection) that caused the event that was observed.

As in this example, the conclusions of many argumentation schemes are considered to be defeasible, and are open to debate by means of *critical questions* associated with each scheme (Walton et al. 2008). For example, one of the critical questions of the above argumentation scheme is whether there is an alternative more plausible explanation for the observed event. Recognition of the argumentation scheme underlying an argument is critical for challenging an argument via critical questions and recognizing answers to those challenges, i.e., in representing and reasoning about scientific debate.

There has been some work on argumentation mining of debate, but none addressing debate in the natural sciences. Teufel et al. (2006) developed a scheme with categories such as support and anti-support for annotating citation function in a corpus of computational linguistics articles. Cabrio and Villata (2012) addressed recognition of support and attack relations between arguments in a corpus of on-line dialogues stating user opinions. Stab and Gurevych (2013) and Stab et al. (2015) are developing guidelines for annotating support-attack relationships between arguments based on a corpus of short persuasive essays written by students and another corpus of 20 full-text articles from the education research domain. Peldszus and Stede (2013) are developing guidelines for annotating relations between arguments which have been applied to the Potsdam Commentary Corpus (Stede 2004). However research on mining debate has not addressed more fine-grained relationships such as asking and responding to particular critical questions of argumentation schemes.

Furthermore, there has been no work on argumentation scheme recognition in scientific text. Feng and Hirst (2011) investigated argumentation scheme recognition using the Araucaria corpus, which contains annotated arguments from newspaper articles, parliamentary records, magazines, and on-line discussion boards (Reed et al. 2010). Taking premises and conclusion as given, Feng and Hirst addressed the problem of recognizing the name of the argumentation scheme for the five most frequently occurring schemes of Walton et al. (2008) in the corpus: *Argument from example* (149), *Argument from cause to effect* (106), *Practical reasoning* (53), *Argument from Consequences* (44), and *Argument from Verbal Classification* (41). (The number of instances of each scheme is given in parentheses.) Classification techniques achieved high accuracy for *Argument from example* and *Practical reasoning*.

Text with genetics content has been the object of study in some previous NLP research. Mizuta et al. (2005) investigated automatic classification of information status of text segments in genetics journal articles. The Colorado Richly Annotated Full Text Corpus (CRAFT) contains 67 full-text articles on the mouse genome that have been linguistically annotated (Verspoor et al. 2012) and annotated with concepts from standard biology ontologies (Bada et al. 2012). The Variome corpus, consisting of 10 journal articles on the relationship of human genetic variation to disease, has been annotated with a set of concepts and relations (Verspoor et al. 2013). None of these corpora have been annotated for argumentation mining.

Finally, Green et al. (2011) identified argumentation schemes in a corpus of letters written by genetic counselors. The argumentation schemes were used by a natural language generation system to generate letters to patients about their case. However, the argumentation in genetics research articles appears more complex than that used in patient communication. Green (2014; 2015) analyzed argumentation in one genetics journal article but did not generalize the results to other articles, nor provide any empirical evaluation.

3 Argumentation Schemes

This section describes ten argumentation schemes that we identified in four research arti-

cles on genetic variants/mutations that may cause human disease (Schrauwen et al. 2012; Baumann et al. 2012; Charlesworth et al. 2012; McInerney et al. 2013). The ten schemes are not the only ones we found but they represent major forms of causal argumentation for the scientific claims of the articles. The schemes are semantically distinct in terms of their premises and conclusions. Most of these schemes are not described in the catalogue of argumentation schemes frequently cited by argumentation mining researchers (Walton et al. 2008). None of them are the same as the ones addressed by Feng and Hirst (2011).

To facilitate comparison of the schemes, they are presented in Table 1 in a standardized format highlighting how the schemes vary in terms of two event variables, X and Y. The articles where the schemes were found are identified by the first initial of the surname of the first author, in parentheses next to the name of the scheme. Most scheme names were chosen for mneumonic value. In the table, X and Y are events, such as the existence of a genetic mutation and disease, respectively, in an individual or group. The premises describe (i) whether the X events have been hypothesized, observed, or eliminated from further consideration, (ii) whether the Y events have been observed, and (iii) (the *Causal potential* column) whether a potential causal relation has been previously established between Xs and Ys. The conclusions, which are understood to be defeasible, describe whether an event X is concluded to have possibly occurred and/or to be a/the possible cause of Y.

As a step towards annotating these argumentation schemes in a corpus, we created initial guidelines containing examples and descriptions of the ten schemes. Illustrating some of the challenges in identifying arguments in this literature, Figure 1 shows the guidelines for two schemes (*Effect to Cause* and *Failed to Observe Effect of Hypothesized Cause*). In Figure 1, three text excerpts from an article are presented. The first two excerpts contain general information needed to interpret the arguments, including one premise of each argument. The third excerpt contains an additional premise of each argument, and conveys the conclusion of each argument. The last sentence of the third excerpt, "He was initially suspected to have EDS VIA, but the urinary LP/HP ratio was within the normal range" conveys the conclusion of the first argument: the patient may have EDS VIA. However, by providing evidence conflicting with that conclusion (the LP/HP data), the sentence also implicit-

ly conveys the conclusion of the second argument: it is not likely that the patient has EDS VIA. The only overt signals of this conflict seem to be the qualifier 'initially suspected' and the 'but' construction.

Our guidelines provide a paraphrase of each argument since many of the example arguments have implicit premises or conclusions (i.e. are enthymemes). For example, the conclusion of the *Failed to Observe Effect of Hypothesized Cause* argument shown in Figure 1 is implicit. In some cases a missing premise is supplied from information presented in the article but not in the given excerpts. In other cases, a missing premise is paraphrased by presenting generally accepted background knowledge of the intended reader such as "A mutation of a gene that is expressed in a human tissue or system may cause an abnormality in that tissue or system." In yet other cases, a conclusion of one argument is an implicit premise of another argument.

As illustrated in Figure 1, each paraphrased argument in the guidelines is followed by a more abstract description of the argumentation scheme. Abstract descriptions of each argumentation scheme are presented in Figures 2 and 3. Note that *Effect to Cause, Eliminate Candidates*, and *Failed to Observe Effect of Hypothesized Cause* are similar but not identical to the *Abductive Argumentation Scheme, Argument from Alternatives*, and *Argument from Falsification* of (Walton et al. 2008), respectively. However, none of the descriptions in (Walton et al. 2008) are attributed to arguments found in the science research literature.

4 Pilot Study

A pilot study was performed to determine the effectiveness of the initial guidelines for identifying a subset of the argumentation schemes. For the study, we added a two-page overview of the task to the beginning of the guidelines and a quiz at the end; and, in the interest of reducing the time required to complete the task, removed the five examples from one article (Charlesworth et al. 2012), which repeated three of the argumentation schemes from the other sources. To summarize, the pilot study materials consisted of examples of eight schemes from (Schrauwen et al. 2012) and (Baumann et al. 2012), and a multiple-choice quiz based upon examples from (McInerney et al. 2013).

The quiz consisted of five multi-part problems, each problem concerning one or more ex-

cerpts from (McInerney et al. 2013) containing an argument. The quiz did not address the task of determining the presence of an argument and its boundaries within an article. Problems I-III tested participants' ability to recognize premises and/or conclusions and names of four key argumentation schemes: *Effect to Cause, Eliminate Candidates, Causal Agreement and Difference*, and *Joint Method of Agreement and Difference*. Problem I (shown in Figure 4) presented a paraphrase of the conclusion of the argument and asked the participant to identify the excerpts containing the premises; and to identify the name of the argumentation scheme from a list of six names (the four scheme names that begin with *Failed* were omitted, although the participant could have selected *None of the above* for those). Problem II asked the participant to select the premises and conclusion of the argument from a list of paraphrases; and to identify the name of the argumentation scheme from the same list of choices given in problem I.

In problem III, the excerpts contained two arguments for the same conclusion. The participant was given a paraphrase of the conclusion and asked to select the excerpt best expressing the premise of the Causal Agreement and Difference argument and the excerpt best expressing the premise of the Joint Method of Agreement and Difference argument. The purpose of problems IV and V was to evaluate participants' ability to interpret more complex argumentation. The excerpts given in problem IV actually conflated multiple arguments. Rather than ask the participant to tease apart the component arguments, the problem asked the participant to select the paraphrases expressing the (main) conclusion and premise. Problem V asked the participant to select the paraphrase best expressing the conclusion of the excerpts in IV and V together.

The study was performed with two different groups of participants. The first group consisted of university students in an introductory genetics class early in the course. They had not received instruction on argumentation in their biology courses, had covered basic genetics in their first two years of study, and had no experience reading genetics research articles. The students were required to participate in the study but were informed that the quiz results would not influence their course grade and that allowing use of their quiz results in our study was voluntary. The students completed the study in 45 to 60 minutes. The results are shown in Table 2.

To comment, since the students had limited relevant background and may not have been motivated to succeed in the task, some of the class did very poorly. The mean number answered correctly on the 11 problems was 49% (N=23). However, six students scored between 73% and 82% (the highest score). The best performance was on Problem I on an *Effect to Cause* argument. This may be due, at least in part, to the fact that this argumentation scheme appeared first in the guidelines and was also the first problem. The question that the fewest number of students answered correctly was III.1, which was to identify the excerpt containing a premise of a Causal Agreement and Difference argument. Overall, the main lesson we learned from this group of study participants, compared to the other participants (see below), is that training and/or motivation need to be improved before running such a study with a similar group of students.

The second group of participants consisted of researchers at several different universities in North America. No compensation was provided. The researchers came from a variety of backgrounds: (A) computer science with no background in genetics, NLP or argumentation, (B) learning sciences with background in argumentation but none in genetics, (C) biology with extensive background in genetics but none in NLP or argumentation, and (D and E) BioNLP researchers. The results are shown in Table 3. Researchers A, C, and D answered all of the questions correctly; B missed only one (III.1); E missed two (II.3 and IV.1). B commented that B did not have sufficient knowledge of genetics to understand the excerpt. The results from this group confirm that several key schemes could be recognized by other researchers based upon reading the guidelines.

5 Annotating Debate

The guidelines do not yet address annotation of relationships between pairs of arguments within an article. Our plan is to annotate the following types of relationships which we found. First, as illustrated by the two arguments shown in Figure 1, two arguments with conflicting conclusions may be presented. Note that four of the argumentation schemes we have identified (see Table 1) may play a prominent role in annotation of this type of relationship, since they provide a way of supporting the negation of the conclusions of other schemes. Second, multiple evidence may be presented to strengthen the premises of an

argument. In the excerpt illustrating *Failed Predicted Effect* in Figure 2, the premise that G is not predicted to have effect P is supported by evidence from three different genetic analysis tools (Mutation Taster, SIFT, or PolyPhen-2).

The third relationship is to preempt an attack by addressing one of the critical questions of an argumentation scheme. One instance of this occurs in (McInerney-Leo et al. 2013), in which a Causal Agreement and Difference argument concludes that a certain variant is the most likely cause of a disease in a certain family, since the occurrence of the variant and the disease is consistent with (the causal mechanism of) autosomal recessive inheritance. Nevertheless, one might ask the critical question whether some other factor could be responsible. Addressing this challenge, a Joint Method of Agreement and Difference argument is given to provide additional support to that claim, since the disease was not found in a control group of individuals who do not have the variant.

6   Conclusion

This paper presented a specification of ten causal argumentation schemes used to make arguments for scientific claims in genetics research journal articles. The specifications and some of the examples from which they were derived were used to create an initial draft of guidelines for annotation of a corpus. The guidelines were evaluated in a pilot study that showed that several key schemes could be recognized by other researchers based upon reading the guidelines.

Acknowledgments

Argumentation Scheme	X	Y	Causal potential	Conclusion
Effect to Cause (B, M)	Unknown	Observed	Yes	X occurred & caused Y
Failed to Observe Effect of Hypothesized Cause (B)	Hypothesized	Not observed	Yes	X did not occur & not cause of Y
Failed to Observe Expected Cause (S)	Not observed	Observed	Yes	X not cause of Y
Consistent with Predicted Effect (S)	Observed	Observed	Yes	X cause of Y
Failed Predicted Effect (C)	Observed	Observed	No	X not cause of Y
Hypothesize Candidates (S, C)	Observed set of Xs	Observed	Yes	One of Xs cause of Y
Eliminate Candidates (S, C, M)	Observed set of Xs but X_0 can be eliminated	Observed	Yes	All Xs but X_0 cause of Y
Causal Agreement and Difference (S, C, M)	Observed in group 1, not in group 2	Observed in group 1, not in group 2	Yes	X cause of Y
Failed Causal Agreement and Difference (C)	Observed in all of group	Not observed in all of group	Yes	X not cause of Y
Joint Method of Agreement and Difference (S, M)	Observed in all of group	Observed in all of group	No	X cause of Y

Table 1. Semantic distinctions among argumentation schemes identified in genetics articles.

I.1	I.2	I.3	II.1	II.2	II.3	III.1	III.2	IV.1	IV.2	V.
16	12	15	9	14	11	4	10	12	13	8

Table 2. Number of students (N=23) who answered each question correctly.

I.1	I.2	I.3	II.1	II.2	II.3	III.1	III.2	IV.1	IV.2	V.
5	5	5	5	5	4	4	5	4	5	5

Table 3. Number of researchers (N=5) who answered each question correctly.

Excerpts from (Baumann et al. 2012):
The Ehlers-Danlos syndrome (EDS) comprises a clinically and genetically heterogeneous group of heritable connective tissue disorders that predominantly affect skin, joints, ligaments, blood vessels, and internal organs … The natural history and mode of inheritance differ among the six major types … Among them, the kyphoscoliotic type of EDS (EDS VIA) … is characterized by severe muscle hypotonia at birth, progressive kyphoscoliosis, marked skin hyperelasticity with widened atrophic scars, and joint hypermobility. …

The underlying defect in EDS VIA is a deficiency of the enzyme lysyl hydroxylase 1 … caused by mutations in *PLOD1* … A deficiency of lysyl hydroxyl results in an abnormal urinary excretion pattern of lysyl pyridinoline (LP) and hydroxylysyl pyridinoline (HP) crosslinks with an increased LP/HP ratio, which is diagnostic for EDS VIA.

At 14 years of age, the index person P1 … was referred to the Department of Paediatrics … for the evaluation of severe kyphoscoliosis, joint hypermobility and muscle weakness. He was initially suspected to have EDS VIA, but the urinary LP/HP ratio was within the normal range.

First Argument Paraphrase:
a. Premise: P1 has severe kyphoscoliosis, joint hypermobility and muscle weakness.
b. Premise: EDS VIA is characterized by severe muscle hypotonia at birth, progressive kyphoscoliosis, marked skin hyperelasticity with widened atrophic scars, and joint hypermobility.
c. Conclusion: P1 may have EDS VIA.

The above is an example of this type of argument:
Effect to Cause (Inference to the Best Explanation)
- Premise (a in example): Certain properties P were observed (such as severe kyphoscoliosis) in an individual.
- Premise (b in example): There is a known potential chain of events linking a certain condition G to observation of P.
- Conclusion (c in example): G may be the cause of P in that individual.

Second Argument Paraphrase:
a. Premise: P1's LP/HP ratio was within normal range.
b. Premise: The underlying defect in EDS VIA is a deficiency of the enzyme lysyl hydroxylase 1 caused by mutations in *PLOD1*. A deficiency of lysyl hydroxyl results in an abnormal urinary excretion pattern of lysyl pyridinoline (LP) and hydroxylysyl pyridinoline (HP) crosslinks with an increased LP/HP ratio.
c. Conclusion: It is not likely that P1 has EDS VIA.

The above is an example of this type of argument:
Failed to Observe Effect of Hypothesized Cause
- Premise (a in example): Certain properties P were not observed (such as increased LP/HP ratio) in an individual.
- Premise (b in example): There is a known potential chain of events linking a certain condition G to observation of P.
- Premise (c in example): G may not be present in that individual.

Figure 1. Description of two argumentation schemes in the initial guidelines.

Effect to Cause
Premise: Certain properties P were observed in an individual.
Premise: There is a potential chain of events linking a condition G to observation of P.
Conclusion: G may be the cause of P in that individual.
Example: See Figure 1.

Failed to Observe Effect of Hypothesized Cause
Premise: Certain properties P were not observed in an individual.
Premise: There is a potential chain of events linking a condition G to observation of P.
Conclusion: G may not be present in that individual (and is not the cause of P in that individual).
Example: See Figure 1.

Failed to Observe Expected Cause
Premise: G is missing from one or more individuals with property P.
Premise: G may be a cause of P in some cases.
Conclusion: G is not likely the cause of P in this case.
Example (Schrauwen): "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."

Consistent with Predicted Effect
Premise: G and P were observed in certain individuals.
Premise: There is a potential chain of events linking G to P.
Conclusion: G may be the cause of certain cases of P.
Example (Schrauwen) : "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 Ca^{2+} influx. This, in turn, could reduce rates of transmitter release and consequently diminish spiral ganglion neuron firing and ascending auditory-pathway activation."
[Note: Earlier in the article, the authors describe finding the c.637+1G>T variant of *CABP2*, which is in the DFNB93 region, in members of a family who were affected with autosomal recessive nonsyndromic hearing loss.]

Failed Predicted Effect
Premise: G was (previously considered to be) a candidate cause of P in some individuals.
Premise: G is not predicted to have effect P.
Conclusion: G may not be a cause of P.
Example (Charlesworth): "The second was a heterozygous missense variant … in exon 2 of *PPM1K* … on chromosome 4. It was not predicted to be damaging by MutationTaster, SIFT, or PolyPhen-2 ..."

Hypothesize Candidates
Premise: There is a potential causal relationship between a certain type of event and a certain type of effect.
Premise: Some individual(s) has experienced a certain effect of that type
Conclusion: There is a set of candidates, one of which may be the cause of that effect.
Example (Charlesworth): "In order to maximize the chances of isolating the causal variant … The first strategy involved selecting only those variants that were present in the exome data of both affected family members for analysis."
[Note: Elsewhere in the article two affected family members whose exome data was analyzed are identified as II-2 and III-7. They are members of a family that exhibits autosomal dominant inheritance of cervical dystonia.]

Figure 2: Some Argumentation Scheme Definitions and Examples

Eliminate Candidates
Premise: There is a set of candidates C, one of which may be the cause of event E.
Premise: Generally accepted statements that explain why some candidates may be eliminated
Premise: One or more members of C can be eliminated as candidates.
Conclusion: One of the remaining members of C may be the cause of E.
Example (Charlesworth): "Homozygous variants, synonymous variants, and variants recorded in dbSNP135 were initially removed. We then filtered out any variant present at a global minor allele frequency (MAF) $\geq$ 1% in a range of publically available databases of sequence variation (1000 Genomes, Complete Genomic 69 Database, and the National Heart, Lung, and Blood Institute [NHLBI] Exome Sequencing Project database), as well as those found in two or more of our own in-house exomes from individuals (n = 200) with unrelated diseases."

Causal Agreement and Difference
Premise: There is a set of individuals I_present that have a feature F and property P.
Premise: There is a set of individuals I_absent that have neither feature F nor property P.
Premise: There is a plausible causal mechanism that could account for the similarities and differences between I-absent and I-present.
Conclusion: F may be the cause of P in I_present.
Example (Charlesworth): "The third, a missense mutation (c.1480A>T …) in exon 15 of *ANO3* … on chromosome 11, segregated perfectly with the disease status in definitely affected and unaffected individuals."

Failed Causal Agreement and Difference
Premise: There is a set of individuals I_present who have feature F and who do not have property P.
Premise: There is a plausible causal link from F to P that could account for the presence of P in I_present if P had occurred.
Conclusion: F may not be a cause of P.
Example (Charlesworth): "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."

Joint Method of Agreement and Difference
Premise: There is a set of individuals I_present that have a feature F and property P.
Premise: There is a set of individuals I_absent that have neither feature F nor property P.
Conclusion: F may be the cause of P in I_present.
Example (Schrauwen): "Next, we checked the inheritance of the *CABP2* variant in the entire Sh10 family (Figure 1) and screened an additional 100 random Iranian controls to ensure that the variant is not a frequent polymorphism. The mutation was not detected in any of the controls..."
[Note: This scheme differs from Causal Agreement and Difference by the absence of the third premise, about a previously known potential causal link from F to P.]

Figure 3: More Argumentation Scheme Definitions and Examples

Problem I.

Excerpts from (McInerney-Leo et al. 2013):
A. Within the ciliopathies, a subgroup of disorders including short-rib polydactyly syndrome (SRPS), Jeune syndrome … are characterized by skeletal abnormalities including a small rib cage, shortening of the long bones, and in some cases, polydactyly … Among the skeletal ciliopathies, the SRPS subtypes are the most severe and are incompatible with postnatal life.
B. To date many mutations causing skeletal ciliopathies affect genes encoding components of the intraflagellar transport (IFT) machinery, a motor-driven trafficking process responsible for transporting proteins required for cilia assembly and function along the axoneme.
C. The first family [SKDP42] is a nonconsanguineous Australian family of predominantly British but also Maori descent, with healthy parents and two affected individuals with SRPS type III.
D. Individual SKDP42.3 presented with short long bones on ultrasound at 16 weeks' gestation. Follow-up ultrasound at 31 weeks demonstrated polyhydramnios, severe shortening of long bones with bowed femurs, macrocephaly, short ribs, and ambiguous genitalia. The baby was born at 32 weeks' gestation but died at 2 hr of age. Autopsy … revealed postaxial polydactyly of both hands ...
E. None of the above excerpts.

1. What evidence was presented that is consistent with the conclusion that the individual referred to as SKDP42.3 (a member of the family identified as SKDP42) had SRPS? (Choose the best single answer from excerpts A-E above.)

2. What general biomedical knowledge explains the connection between the evidence you selected (in A-E) and the diagnosis that individual SKDP42.3 had SRPS? (Choose the best single answer from excerpts A-E above.)

3. Select the argumentation scheme that best fits the argument for the diagnosis (Choose the best single answer from the following):
   • Causal Agreement and Difference
   • Consistent with Predicted Effect
   • Effect to Cause (Inference to the Best Explanation)
   • Eliminate Candidates
   • Hypothesize Candidates
   • Joint Method of Agreement and Difference
   • None of the above.

Figure 4. Part of quiz used in pilot study. The answers are 1-D, 2-A, and 3-Effect to Cause

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". *Am J Hum Genetics* 90, 201-216

Cabrio, E. and Villata, S. 2012. Generating Abstract Arguments: A Natural Language Approach. In Verheij, B., Szeider, S., and Woltran, S. (eds.) *Computational Models of Argument: Proceedings of COMMA 2012*. Amsterdam, IOS Press, 454-461.

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

Cohen, K.B. and Demner-Fushman, D. 2014 *Biomedical Natural Language Processing*, Amsterdam: John Benjamins Publishing Company

Feng, V.W. and Hirst, G. 2011. Classifying Arguments by Scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,* Portland, OR, 987-996.

Green, N., Dwight, R., Navoraphan, K., and Stadler, B. 2011. Natural Language Generation of Transparent Arguments for Lay Audiences. *Argument and Computation* 2(1): 23-50.

Green, N. L. 2014. Towards Creation of a Corpus for Argumentation Mining the Biomedical Genetics Research Literature. In *Proc. of the First Workshop on Argumentation Mining*, ACL 2014, Baltimore.

Green, N. L. 2015. Argumentation for Scientific Claims in a Biomedical Research Article. In Cabrio, E., Viallata, S., and Wyner, A. (Eds.) *ArgNLP 2014: Frontiers and Connections between Argumentation Theory and Natural Language Processing*, Forli-Cesena, Italy, July 21-25, 2014, CEUR Workshop Proceedings, Vol-1341.

Liakata, M, et al. 2012. Automatic Recognition of Conceptualization Zones in Scientific Articles and Two Life Science Applications. *Bioinformatics* 28(7).

McInerney-Leo, A.M. et al. 2013. Short-Rib Polydactyly and Jeune Syndromes Are Caused by Mutations in WDR60. *Am J Hum Gen* 93, 515-523, Sept 5, 2013

Mizuta, Y., Korhonen, A., Mullen, T. and Collier, N. 2005. Zone Analysis in Biology Articles as a Basis for Information Extraction. *International Journal of Medical Informatics* 75(6): 468-487.

Mochales, R. and Moens, M. 2011. Argumentation Mining. *Artificial Intelligence and Law* 19, 1-22.

Peldszus, A. and Stede, M. 2013. From Argument Diagrams to Argumentation Mining in Texts: A Survey. *Int. J of Cognitive Informatics and Natural Intelligence* 7(1), 1-31)

Reed, C., Mochales-Palau, R., Moens, M., and Milward, D. 2010. Language Resources for Studying Argument. In *Proceedings of the 6th Conference on Language Resources and Evaluation, LREC2008, ELRA*, 91-100.

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.

Stab, C. and Gurevych, I. 2014. Annotating Argument Components and Relations in Persuasive Essays. In *Proc. COLING 2014*, pp. 1501-1510

Stab, C., Kirschner, C., Eckle-Kohler, J., and Gurevych, I. 2015. Argumentation Mining in Persuasive Essays and Scientific Articles from the Discourse Structure Perspective. In Cabrio, E., Viallata, S., and Wyner, A. (Eds.) *ArgNLP 2014: Frontiers and Connections between Argumentation Theory and Natural Language Processing*, Forli-Cesena, Italy, July 21-25, 2014, CEUR Workshop Proceedings, Vol-1341

Stede, M. 2004. The Potsdam Commentary Corpus. In *Proc. ACL Workshop on Discourse Annotation,* Barcelona, Spain, pp. 96-102.

Teufel, S. 2010. The Structure of Scientific Articles: Applications to Citation Indexing and Summarization. Stanford, CA, CSLI Publications.

Teufel, S., Siddharthan, A., and Tidhar, D. 2006. An Annotation Scheme for Citation Function. In *Proc. of SIGDIAL-06*, Sydney, Australia.

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

Verspoor, K., Yepes, A.J., Cavedon, L., et al. 2013. Annotating the Biomedical Literature for the Human Variome. *Database*, Vol. 2013, Article ID bat019

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