TREC 2006 Genomics Track Guidelines for Relevance Judges

The goal of the TREC Genomics Track is to improve information retrieval in the biomedical domain. Each year, 30-40 research groups develop algorithms that automatically interpret questions collected from real biologists into queries that can be applied against databases of scientific articles. The groups submit their best search results, which are pooled and judged for relevance to the original question. This year, 160,000 full-text articles from 59 journals will be searched. See http://ir.ohsu.edu/genomics/2006protocol.html for more details.

The task for this year has 28 questions that fall into four categories based on the information sought. These are called General Topic Types, or GTTs, shown in Table 1. Your job as a relevance judge is to assess the relevance of submitted paragraphs, isolate the minimum information in the paragraph that answers the question, and group relevant answers by similarity.

Submitted paragraphs will be judged as being definitely relevant, possibly relevant, or not relevant according to the guidelines presented below.

Table 1. General Topic Types and Examples.

GTT	Question Pattern	Example
Find articles describing the role of a gene involved in a given disease.	What is the role of gene in disease?	What is the role of DRD4 in alcoholism?
Find articles describing the role of a gene in a specific biological process.	What effect does gene have on biological process?	What effect does the insulin receptor gene have on tumorigenesis?
Find articles describing interactions (e.g., promote, suppress, inhibit, etc.) between two or more genes in the function of an organ or in a disease.	How do genes interact in organ function?	How do HMG and HMGB1 interact in hepatitis?
Find articles describing one or more mutations of a given gene and its biological impact.	How does a mutation in gene influence biological process?	How does a mutation in Ret influence thyroid function?

Instructions

Upon receipt of pooled results, first review the question. If there is a gene or protein mentioned, identify synonyms for it. For biological processes or diseases, familiarize

yourself with more general concepts, as well as sub-topics (see Appendix I for suggested resources). For example, “mad cow disease” in Topic 160 is formally known as bovine spongiform encephalopathy, abbreviated as BSE. It is a member of the Transmissable Spongiform Encephalopathies (TSE) disease family, of which Creutzfeld-Jacob disease (CJD) is another member (these relationships may be found by looking up the disease names in MeSH, described below). Therefore, definitely relevant passages refer to mad cow or BSE, possibly relevant passages refer to the TSE family, of which mad cow disease is a member. References to the related, but different disease, CJD, are not relevant. In another example, Topic 179 asks about liver function. The definition of “liver development” in the Gene Ontology states that the liver secretes bile, synthesizes blood clotting factors and vitamin A, and stores glycogen. Those functions that are unique to the liver and are supplied as an answer are definitely relevant. References to functions that occur in other organs in addition to the liver are possibly relevant.

> TIP: Before you begin to judge, scan the results and look for frequently appearing terms. Look them up to see how they relate to the concepts in the question. Are they sub-topics or synonyms?

1. Evaluate paragraphs for relevance

You will be presented with paragraphs from full-text articles. These paragraphs contain the portions of text submitted by the participants. Read through the paragraph to determine if it is definitely relevant, possibly relevant, or not relevant to the question. Table 2 shows examples of judgments for six of the questions in this year’s event. In general, a paragraph is definitely relevant if it contains all required elements of the question AND it answers the question. A paragraph is possibly relevant if it contains the majority of required elements, missing elements are within the realm of possibility (more general terms are mentioned that probably include the missing elements), AND it possibly answers the question.

Table 2. Examples and Reasoning for Relevance Judgments

Topic ID	Question	Excerpt	Judgment	Reason
160	What is the role of PrnP in mad cow disease?	Bovine Prion Protein Gene (PRNP) Promoter Polymorphisms Modulate PRNP Expression and May Be Responsible for Differences in Bovine Spongiform Encephalopathy Susceptibility	Definitely Relevant	
160	What is the role of PrnP in mad cow disease?	Transmissible spongiform encephalopathies (TSEs), or prion diseases, are mammalian neurodegenerative disorders characterized by a posttranslational conversion and brain accumulation of an insoluble, protease-resistant isoform (PrPSc) of the host-encoded cellular prion protein (PrPC)	Possibly Relevant	Refers to family of disorders (“TSEs”) that includes mad cow disease. Not clear whether findings apply to mad cow disease.

160	What is the role of PrnP in mad cow disease?	The central role of PrP in the transmissible spongiform encephalopathies (TSEs), the proximity of the gene which encodes doppel (Prnd) to the PrP gene (Prnp) and the structural similarity shared by PrP and doppel have led to the proposition that ataxia which develops during TSE disease could, in part, be due to doppel.	Not Relevant	Excerpt discusses role of gene adjacent to PrnP
161	What is the role of IDE in Alzheimer's disease?	Taken together these results suggest that the use of insulysin to hydrolyze A peptides represents an alternative gene therapeutic approach to the treatment of Alzheimer's disease.	Definitely Relevant	IDE is the gene symbol for the protein insulysin
161	What is the role of IDE in Alzheimer's disease?	there is an inverse correlation between in vivo insulysin activity levels and brain A peptide levels and suggest that modulation of insulysin activity may alter the risk for Alzheimer's disease.	Definitely Relevant	IDE is the gene symbol for the protein insulysin
169	How does APC protein affect actin assembly?	we showed that APC clusters were colocalized with DLG protein at cellular protrusions of subconfluent MDCK cells. A portion of the clusters was found at the tips of microtubules extending into the cellular protrusions. In addition, actin stress fibers converged near the clusters.	Possibly relevant	APC not shown to affect actin assembly, just co-localize with it.
174	How does BRCA1 ubiquitinating activity contribute to cancer?	Both genes contribute to homologous recombination and DNA repair, to embryonic proliferation, to transcriptional regulation and, for BRCA1, to ubiquitination. But questions regarding BRCA1 and BRCA2 biology remain, and their resolution is critical for clinical development. Why do ubiquitously expressed genes that participate in universal pathways lead, when mutant, specifically to breast and ovarian cancer?	Definitely Relevant	
174	How does	The RING heterodimer BRCA1-BARD1	Definitely	

	BRCA1 ubiquitinating activity contribute to cancer?	is a ubiquitin ligase inactivated by a breast cancer-derived mutation	Relevant	
174	How does BRCA1 ubiquitinating activity contribute to cancer?	Defining biochemical functions for the BRCA1 tumor suppressor protein: analysis of the BRCA1 binding protein BAP1	Not Relevant	Biochemical functions could refer to ubiquitinating activity, but it is too vague to count as relevant.
174	How does BRCA1 ubiquitinating activity contribute to cancer?	Proteasome-mediated degradation of BRCA1 protein in MCF-7 human breast cancer cells.	Possibly Relevant	Proteasome is part of the ubiquitin pathway, but not limited to it.
174	How does BRCA1 ubiquitinating activity contribute to cancer?	When monoubiquitinated, the FANCD2 protein co-localizes with the breast cancer susceptibility protein BRCA1 in DNA damage induced foci.	Possibly Relevant	Link between ubiquitin on FANCD2 and BRCA1 is not specified.
174	How does BRCA1 ubiquitinating activity contribute to cancer?	The recent identification of the ubiquitin protein ligase activity of BRCA1 implies a possible functional connection between both genes.	Not Relevant	No mention of cancer.
176	What interactions between CFTR and Sec61 cause degradation of CFTR, leading to cystic fibrosis?	We show that the level of the Sec61.CFTR complexes are highest when CFTR degradation proceeds at the greatest rate (approximately 90 min after pulse labeling).	Definitely Relevant	
176	What interactions between CFTR and Sec61 cause degradation of CFTR, leading to cystic fibrosis?	CFTR expression and ER-associated degradation in yeast.	Possibly Relevant	Sec61 may be part of ER-associated degradation apparatus.
176	What interactions between CFTR and Sec61 cause	Many cystic fibrosis transmembrane conductance regulator (CFTR) mutants are recognized as aberrant by the quality control apparatus at the endoplasmic reticulum (ER) and are	Possibly Relevant	Sec61 may be part of quality control apparatus.

	degradation of CFTR, leading to cystic fibrosis?	targeted for degradation. Replacement of arginine residues at positions R29, R516, R555, and R766 with lysine residues to inactivate four of these motifs simultaneously causes delta F508 CFTR, present in approximately 90% of CF patients, to escape ER quality control and function at the cell surface.		
176	What interactions between CFTR and Sec61 cause degradation of CFTR, leading to cystic fibrosis?	We now have identified three short oligopeptide regions in the C-terminal domain which impact cystic fibrosis transmembrane conductance regulator (CFTR) maturation and stability in different ways	Not Relevant	No mention of possible degradation pathway.
182	How do mutations in Sonic Hedgehog genes affect developmental disorders?	No SHH mutation was found in six polymalformed cases combining HPE with other defects, such as skeletal, limb, cardiac, anal and/or renal anomalies.	Definitely Relevant	Even though no mutations in SHH were found associated with the indicated disorders, the gene was sufficiently screened to allow a definitive conclusion about it's role to be made.
182	How do mutations in Sonic Hedgehog genes affect developmental disorders?	Sonic hedgehog (Shh) has been proposed to function as an inductive and trophic signal that controls development of epaxial musculature in vertebrate embryos.	Not Relevant	No mutations or disorders mentioned.

There are some questions for which only a few paragraphs are definitely relevant or possibly relevant. In past years, 50-95% of submitted results were not relevant. It is tempting to show leniency when few relevant paragraphs are encountered. Resist the urge to relax criteria for relevance and try to maintain consistent evaluation standards. Judging consistency is also affected by familiarity with the topic. As you judge results, you will become more familiar with the area of research, and may change your criteria for relevance.

TIP: After completing judgments for a question, return to the first 10% of results and determine whether the same relevance criteria were maintained throughout.

2. Identify the minimum portion of text required to answer the question without having to refer back to the full-text for additional information

Table 3 shows some examples of adequate and inadequate excerpts. Pronouns (e.g., these, they, it) that do not reference the relevant subject within the extracted text are inadequate. Reference to the relevant subject as an unspecified generic component (e.g., "the subunit", "these genes", "this disease") defined within the extracted text is acceptable. However, these are not acceptable if not defined within the extract. Acronyms or abbreviations that are explained elsewhere but whose definitions are not part of the excerpt are acceptable, as these are considered to be functioning as synonyms. Minimum excerpts may range from a portion of a sentence to the entire paragraph, but they must contain all required elements from the question to be definitely relevant.

Table 3. Minimum Excerpt Evaluation

Topic ID	Question	Excerpt	Judgment	Reason
176	What interactions between CFTR and Sec61 cause degradation of CFTR, leading to cystic fibrosis?	Finally a mutation (sec61-2) in the translocon protein Sec61p that prevents retrotranslocation across the ER membrane also blocked degradation.	Inadequate	Doesn't specify CFTR
176	What interactions between CFTR and Sec61 cause degradation of CFTR, leading to cystic fibrosis?	This requires retrograde translocation of proteins from the ER back to the cytoplasm, which is mediated by Sec61, the central component of the ER protein-import channel. Underactive or overactive ER degradation machinery contributes to the pathogenesis of several severe human diseases.	Inadequate	Doesn't specify CFTR
160	What is the role of PrnP in mad cow disease?	Animals devoid of the gene encoding PrP (Prnp-deficient) do not develop disease after transmission (3-5) and in transgenic mice with higher copy numbers of Prnp, progression of the disease is faster.	Inadequate	Unspecfied "disease" is defined earlier in the text.
161	What is the role of IDE in Alzheimer's disease?	While the unequivocal effects of these genes in the pathogenesis of AD await the identification of all of the functional variants and testing them in large series, these findings provide evidence for the existence of LOAD risk variants that	Indequate	"These genes" refers to three genes mentioned earlier in the paragraph.

		affect Aß, within and/or in the vicinity of these genes.		
161	What is the role of IDE in Alzheimer's disease?	IDE is located on chromosome 10q23.3 close to a region of linkage for LOAD.	Adequate	"LOAD" is defined earlier in the text as Late Onset Alzheimer Disease

3. Code results with MeSH terms

Relevant extracts that come from different articles may contain largely the same information (e.g. research that is frequently cited in the introduction of subsequent articles). Participants in TREC 2006 will be rewarded for the breadth of answers submitted. To determine breadth, excerpts will be coded with standardized terms, allowing grouping of related results. The standardized terms are taken from the Medical Subject Headings (MeSH) thesaurus, developed and used by the National Library of Medicine to index articles in Medline (http://www.nlm.nih.gov/mesh/meshhome.html). PubMed uses MeSH terms to automatically expand keywords entered into the search box, usually resulting in higher quality and quantity of search results.

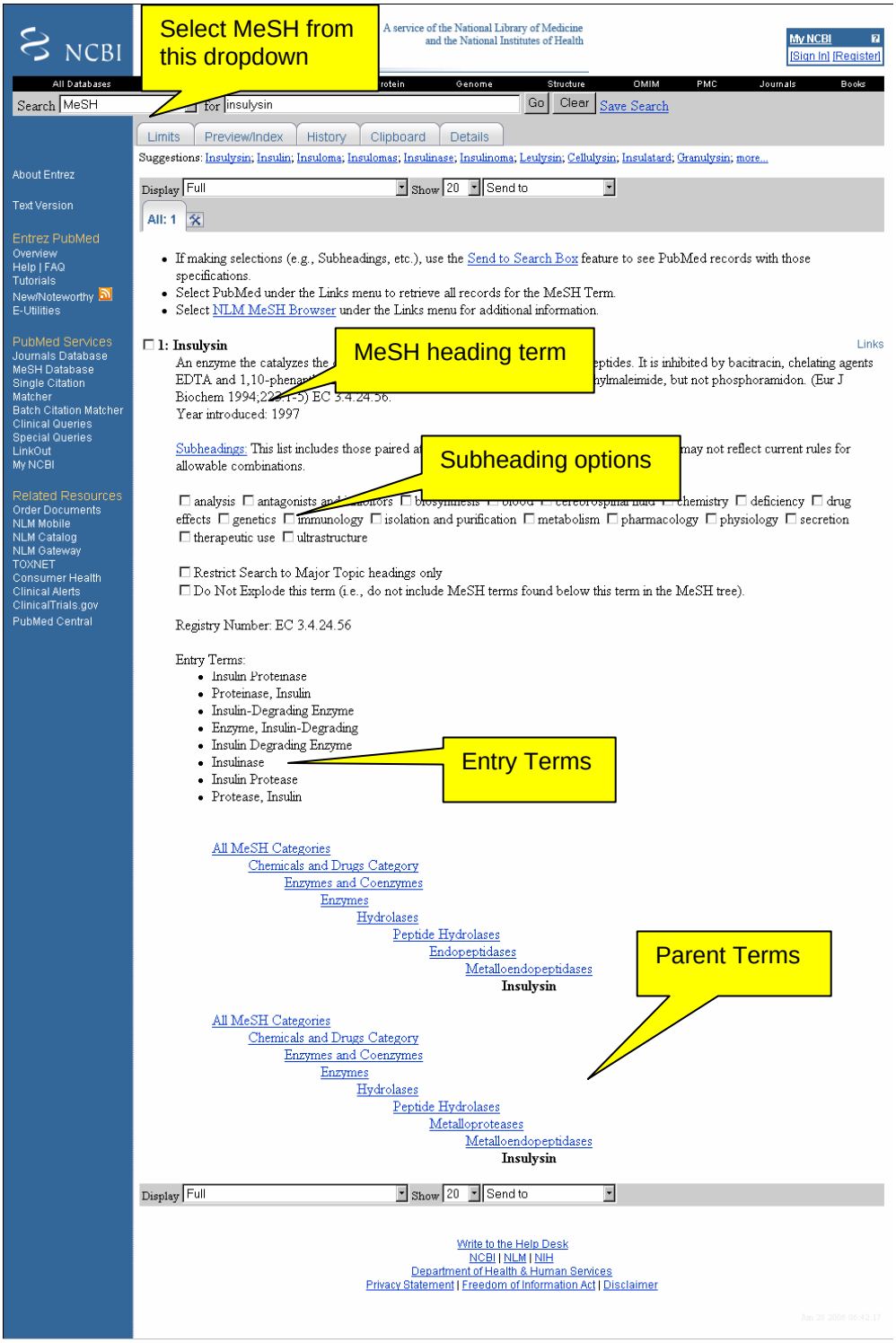
MeSH has different types of terms, including headings, subheadings, and entry terms (Figure 1). Text extracts should be labeled with individual MeSH headings, without subheadings if possible. A combination of a MeSH heading and subheading can be used if required. An extracted passage may be tagged with more than one MeSH term if necessary. Entry terms should only be used to assist with searching MeSH, and not used in assignments.

When selecting MeSH terms, look for terms that contribute to answering the question. Terms that merely replicate one of the required elements of the question aren't useful for distinguishing different types of answers. Sometimes the answer to a question is best represented by an action term, such as "peptide hydrolysis". Nouns describing objects (versus actions) are more prevalent in MeSH, and the MeSH term "peptide hydrolases" should be considered as an alternative.

You can search for MeSH terms in two ways:
1. Use the MeSH Browser (http://www.nlm.nih.gov/mesh/MBrowser.html) to find appropriate terms..
2. Search MeSH from the PubMed interface (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) (shown in Figure 1). In the upper left corner, change the drop-down menu from "PubMed" to "MeSH". When you search MeSH, you will be presented with a set of results consisting of the official MeSH term, and a brief description. Clicking on the term opens the full record, which shows all the synonyms linked to the official term. You will enter the official MeSH term. Table 4 shows some terms applied to the relevant excerpts from Table 2. Although many MeSH

terms may apply, select as few as possible that distinguish one set of related excerpts from another.

Figure 1. Searching MeSH from the PubMed Browser.



Note in the examples from Topic 161 in Table 4 that the MeSH term selected is “Peptide Hydrolases”. This term was found by looking up “IDE” in iHoP (see Appendix I), which shows “insulysin” as a synonym. Looking up “insulysin” in MeSH shows that it has the parent term metalloendopeptidase, which has the parent term peptide hydrolase (see Figure 1). The excerpts talk about insulysin hydrolyzing A-beta peptides, but there is no reference to metalloendopeptidase. Therefore, the MeSH term “peptide hydrolases” links the enzymatic function mentioned in two passages to the gene symbol used in the original question.

Initially, judges should err on the side of assigning more MeSH terms than absolutely required, rather than less. A secondary review by another judge will be applied to condense the set of MeSH terms applied to a topic question to the minimum set required to cover the main aspects of the answers to the topic question.

Table 4. MeSH Terms Applied to Excerpts

Topic ID	Question	Excerpt	MeSH term	Reason
160	What is the role of PrnP in mad cow disease?	Bovine Prion Protein Gene (PRNP) Promoter Polymorphisms Modulate PRNP Expression and May Be Responsible for Differences in Bovine Spongiform Encephalopathy Susceptibility	Polymorphism, Genetic	PRNP polymorphisms influence mad cow disease.
160	What is the role of PrnP in mad cow disease?	Transmissible spongiform encephalopathies (TSEs), or prion diseases, are mammalian neurodegenerative disorders characterized by a posttranslational conversion and brain accumulation of an insoluble, protease-resistant isoform (PrPSc) of the host-encoded cellular prion protein (PrPC)	PrPSc Proteins	PrPSc is the general name of the disease-forming isoform of prion proteins like PrnP.
174	How does BRCA1 ubiquitinating activity contribute to cancer?	Both genes contribute to homologous recombination and DNA repair, to embryonic proliferation, to transcriptional regulation and, for BRCA1, to ubiquitination. But questions regarding BRCA1 and BRCA2 biology remain, and their resolution is critical for clinical development. Why do ubiquitously expressed genes that participate in universal pathways lead, when mutant, specifically to breast and ovarian cancer?	Breast Neoplasms	Specifies type of cancer. No MeSH term for “ubiquitination” found.
174	How does BRCA1 ubiquitinating activity contribute to	The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation	Ubiquitin-Protein Ligase Complexes, Breast Neoplasms	Specifies role in ubiquitin pathway; specified type of cancer.

	cancer?			
161	What is the role of IDE in Alzheimer's disease?	Taken together these results suggest that the use of insulysin to hydrolyze A peptides represents an alternative gene therapeutic approach to the treatment of Alzheimer's disease.	Amyloid beta-Protein Precursor, Peptide Hydrolases	It is the peptidase activity of insulysin acting on A beta that is thought to impact Alzheimer's disease.
161	What is the role of IDE in Alzheimer's disease?	there is an inverse correlation between in vivo insulysin activity levels and brain A beta peptide levels and suggest that modulation of insulysin activity may alter the risk for Alzheimer's disease.	Amyloid beta-Protein Precursor, Peptide Hydrolases	It is the peptidase activity of insulysin acting on A beta that is thought to impact Alzheimer's disease.

Appendix I. Resources for Judging.

IHoP – Information Hyperlinked over Proteins
http://www.ihop-net.org/UniPub/iHOP/
This database lists synonyms for proteins and provides excerpts from the literature, allowing you to familiarize yourself with the biology of the protein.

PubMed Books
http://www.ncbi.nlm.nih.gov/entrez/query/Books.live/Help/bookhelp.html#search
Good for general overviews of biological processes and diseases. Link takes you to instructions for searching books.

AmiGO, the Gene Ontology browser
http://www.godatabase.org/cgi-bin/amigo/go.cgi
Good for brief definitions of biological processes. No disease information. Use MeSH or PubMed books.

MeSH – Medical Subject Headings
http://www.nlm.nih.gov/mesh/MBrowser.html
For biological processes and diseases, provides synonyms or constituent processes that are part of the indicated concept.