Report on the TREC 2003 Experiment: Genomic Track

Patrick Ruch[1ab], **Christine Chichester**[c], **Gilles Cohen**[a], **Giovanni Coray**[b]
Frédéric Ehrler[ad], **Hatem Ghorbel**[b], **Henning Müller**[a], **Vincenzo Pallotta**[b]

[a]*SIM, University Hospital of Geneva, Geneva*
[b]*Theoretical Computing Laboratory, Swiss Federal Institute of Technology, Lausanne*
[c]*GeneBio SA, Geneva*
[d]*AI Laboratory, University of Geneva, Geneva*

Summary

For our first participation to the TREC evaluation campaign, our efforts concentrated on the genomic track. Because we joined the competition at the end of June, we were not able to submit runs for the ad hoc retrieval task (task I), and therefore this report mostly focuses on the information extraction task (task II).

Task I. Our approach uses thesaural resources (from the UMLS) together with a variant of the Porter stemmer for string normalization. Gene and Protein Entities (GPE) of the collection (525,938 MedLine citations) were simply marked up by dictionary look up during the indexing in order to avoid erroneous conflation: strings not found in the UMLS Specialist lexicon (augmented with various English lexical resources) were considered as GPE and were moderately overweighted. In the same spirit like other TREC competitors [23] for task I, an overweighting factor was also applied to features belonging to Medical Subject Headings (MeSH) and found in MedLine citations using a MeSH mapping tool [1]. A standard vector space IR engine with tf-idf parameters was used for indexing the Genomic collection: article's titles, MeSH and RN terms, and abstact fields were selected. Best average precisions were obtained with atc.ntn (using the SMART notation) schemes: 16.71 (standard) vs. 17.02 (using UMLS resources and GPE tagging). Studies made after the competition and inspired by results reported by other groups [13][14] confirmed that narrowing the search to those species appearing in the query provides a very effective improvement of the average precision. The species are detected based on their listing in a dictionnary (extracted from the MeSH terminology). The refinement strategy consists in filtering out documents when the targetted species is not found in the abstract. After retrieval, this simple strategy yield to an important improvement of the average precision: from 17.02 up to 35.80.

Task II. Our approach is based on argumentative structuring, i.e. classification of textual segments into argumentative classes. We see the task as a question-answering task using always the very same question. We take advantage of a classifier likely to predict the argumentative class of any textual segment with high accuracy (F-score = 85%). We observe that when not taken from the title, GeneRIFs are found -ranked by decreasing order- in: 1) CONCLUSION, 2) PURPOSE, 3) RESULTS, 4) METHODS. After sentence splitting, sentences are classified and ranked according to these four categories. On that basis, a second ranking is made based on the similarity with the title (45% of GeneRIFs; Dice baseline = 50.47%). Then, we compute a combined score for each of these features, setting a Dice-like threshold to decide whether we use the title or the best scored sentence as GeneRIF. Finally, a last step consists in narrowing segment boundaries to shorten the length of the candidate GeneRIF. A set of ad hoc and argumentative filters are applied in order to remove irrelevant pieces at the end/beginning of the selected segment. Examples of phrases that are removed at the beginning are "in this paper, finally...". Then, sentence endings (up to 7 words) classified as METHODS (such as "in contrast to current models", "by the...") are also removed. Using the complete article instead of the abstract did not result in any improvement. Our best performances are obtained by using 14 (Dice = 52.78%) and 23 (Dice = 52.41%) segments from the abstract, while the reamining originates from the title. The use of argumentative features is encouraging, however more complex features combination will have to be explored in the future.

Introduction

Systems for text mining are becoming increasingly important in biomedicine because of the exponential growth of knowledge. The mass of scientific literature needs to be filtered and categorized to provide for the most efficient use of the data. The problem of accessing this increasing volume

[1] Contact author. Email: patrick.ruch@epfl.ch

of data demands the development of systems that first, can retrieve pertinent information from unstructured texts and second, can help professional curators to annotate high-quality DBs in the biomedical domain (as in SwissProt with Gene Ontology annotations [7] [24] or in MedLine with MeSH annotations [2]). The former task as been largely addressed in previous TREC studies, at least from a general point of view, however it is the first time TREC investigates *ad hoc* retrieval in genomics. The second task of the TREC 2003 Genomic track aims at extracting the Gene Reference Into Function (GeneRIF), as provided in LocusLink, within a corpus of MedLine citations. For this last task, full-text articles are also available.

Input

Locus - ABCA1: ATP-binding cassette, sub-family A (ABC1), member 1

MedLine record - PMID - 12804586

TI - Dynamic regulation of alternative ATP-binding cassette transporter A1 transcripts.

AB – […] The longest (class 1) transcripts were abundant in adult brain and fetal tissues. Class 2 transcripts predominated in most other tissues. The shortest (class 3) transcripts were present mainly in adult liver and lung. To study the biochemical significance of changes in transcript distribution, two cell models were compared. In primary human fibroblasts, upregulation of mRNA levels by oxysterols and retinoic acid increased the relative proportion of class 2 transcript compared to class 1. Phorbol ester stimulated human macrophage-derived THP-1 cells increased the abundance of class 1 transcripts relative to class 2. In both cell lines class 3 transcript levels were minimal and unchanged. It is shown here for the first time that the **regulation of ABCA1 mRNA levels exploits the use of alternative transcription start sites**.

Output

GeneRIF - **regulation of ABCA1 mRNA levels exploits the use of alternative transcription start sites**

Figure 1. Example of a record in LocusLink and the corresponding GeneRIF.

In order to evaluate our hypothesis, we uses the *easyIR* system (http://lithwww.epfl.ch/~ruch/softs/softs.html), which implements standard vector space IR schemes. The extraction of the gene function is seen as a sentence selection task [3][9][10][11][12] and is conducted using an argumentative classifier (called *LASt*, cf. the same link). Our experiments were conducted on an Intel Pentium IV/2.0, with 2GB of memory and 2 x 240 GB of (external USB-2) disks[2]. All experiments were fully automatic.

Background

In order to have an overall view of the underlying problems in generating the most appropriate GeneRIF during the last TREC genomic evaluation campaign [9], an example of a record in LocusLink is given in Figure 1. In the top part of this figure, we find the locus ("ABCA1") and the MedLine record identifier ("PMID – 12804586"). After the label "TI", we have reported the article's title and the abstract is given after the label "AB". In this case, we can see that the corresponding GeneRIF is extracted from the abstract. A typical GeneRIF extraction task is defined as follows: given a PMID (a PubMed reference to a MedLine citation), find the function of a given gene (Figure 1).

Preliminary studies [19] [20] showed that around 95% of the GeneRIF snippets are extracted from the title or from the abstract of the linked scientific paper. Moreover, from this set, around 42% were direct "cut and paste" from either the title or the abstract (Figure 1 is such an example) while another 25% contained significant portions of the title or abstract.

Latent Argumentative Structuring

In MedLine citations, abstracts are sometimes provided with explicit argumentative moves, such as "BACKGROUND", "AIM AND BACKGROUND", "PURPOSE", "METHODS", "RESULTS", "DISCUSSION", "CONCLUSION"… Unfortunately these explicit structural markers are neither stable, nor mandatory; therefore it is difficult to rely on such explicit markers. Although the labels that are used to express these moves are unstable, the hypothesis supporting this study is that conclusion sentences would be good candidates for identifying key/novelty facts in scientific texts; thus supporting gene functions in genomic corpora. Indeed, as stated in professional guidelines (ANSI/NISO Z39. 14-1979), articles in experimental sciences tend to respect strict argumentative patterns with at least 4 sections: PURPOSE-METHODS-RESULTS-CONCLUSION. Several studies confirm that at least the above 4 moves –leaving aside minor variation of labels– are reported to be very stable across different scientific genres (chemistry, anthropology, computer sciences, linguistics…) [4], and are confirmed in biomedical [5] [6].

[2] The indexing of the MedLine collection took more than 250 hours, and we encountered some problems with the first indexing, so that we had to run the process twice.

With position of sentences				
	PURP	METH	RESU	CONC
PURP	80.65%	0%	3.23%	**16**%
METH	10%	70%	10%	10%
RESU	18.58%	5.31%	23.89%	52.21%
CONC	**18.18**%	0%	2.27%	79.55%
Without position of sentences				
	PURP	METH	RESU	CONC
PURP	93.55%	0%	3.23%	**3**%
METH	30%	70%	0%	0%
RESU	27.43%	5.31%	23.01%	44.25%
CONC	**2.27**%	0%	2.27%	95.45%

Table 1. Confusion matrices for the argumentative classifier. The position is useful to separate between PURPOSES and CONCLUSION classes.

In table 1, we give the confusion matrix of the argumentative classifier. The F-score for the overall classification task is about 85%, but important variations are observed depending on the considered binary classification: if CONCLUSION and PURPOSES classes are well classified, the RESULTS class is mostly ill defined and cannot be accurately separated from CONCLUSION.

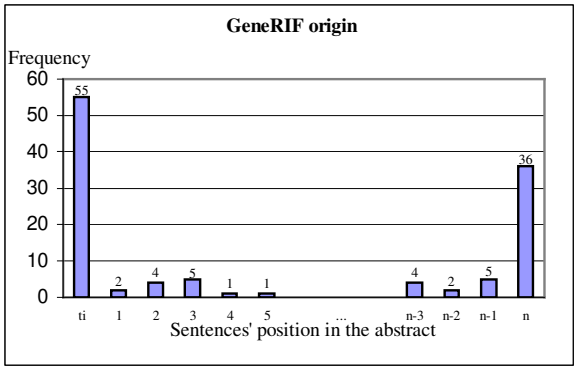

Figure 2. Distribution[3] of the GeneRIF position in the title (ti) and abstract.

We have also analyzed the distribution of sentences location used to produce the GeneRIF. In this case, we consider together the title (the first column labeled "ti" in Figure 2) and sentences from the abstract. From the 139 GeneRIFs used in our experiments, 55 are extracted from the title as depicted in Figure 2. The second most frequent source of GeneRIF is the last sentence of the abstract (last column of Figure 2 with the label "n") which provides 36 geneRIFs. Between these two extreme positions, the distribution of the GeneRIF location is rather flat. It is interesting to observe that the argumentative distribution (Table 3) does not fully match the positional ditribution in the abstract, but the two distributions tend to correlate (Table 3).

[3] Courtesy of Jacques Savoy.

Genomics Information Extraction

A sentence splitting module was designed for the task in order to take into account specific and frequent usages of the dot character, such as in decimals and acronyms. In table 2, we separate between GeneRIFs found by using only the title of the abstract and other GeneRIFs.

Origin	Number	Query ID
Title	9	9, 10, 16, 74, 88, 123, 126, 133
Other	131	Other queries

Table 2: Distribution between GeneRIFs found in titles (with classic Dice = 100%, when using the title) and those found elsewhere.

To identify where human GeneRIFs originates from, regarding argumentative criteria, we apply the LASt system to analyse the distribution of 139 (i.e. the complete data set) and 131 (i.e. excluding queries, whose GeneRIF originates from the title) GeneRIFs into each of the argumentative classes: results are reported in table 3. A sample of the output data in given at the end of the report in Annex 1.

CONCLUSION	72 (51.8%)	68 (51.9%)
PURPOSE	59 (43.9%)	58 (44.3%)
RESULTS	3 (2.2%)	3 (2.3%)
METHODS	3 (2.2%)	2 (1.5%)
Total	139 (100%)	131 (100%)

Table 3. Distribution of argumentative classes among GeneRIFs.

These results are consistent with the confusion matrix given in Table 1: PURPOSES and CONCLUSION hard hardly separable regarding strict lexical information (confusion between 16% amd 18.18% in table 1), therefore positional information becomes important. This information is already combined in the LASt classification and time was too short to redesign the classifier. Unlike for argumentative classification, it is observed that confusion between RESULTS and CONCLUSION classes is not a major issue for GeneRIF extraction.

Combining argumentative classes with titles

Unfortunately, our first submitted run (run 0) computed by selecting the best CONCLUSION sentence as GeneRIF results in performances below the baseline, as shown in Table 4. The baseline, in Table 5, is calculated by simply selecting the title of the abstract and we can notice that more than half of the submitted runs were below this baseline.

Classic Dice	35.20%
Modified unigram Dice	34.57%
Modified bigram Dice	20.04%
Modified bigram Dice phrases	21.58%

Table 4. Results when selecting GeneRIFs considering only sentences classified as CONCLUSION (Run 0).

Classic Dice[4] (Dice1)	50.47%
Modified unigram Dice (Dice2)	52.60%
Modified bigram Dice (Dice3)	34.82%
Modified bigram Dice phrases (Dice4)	37.91%

Table 5. Baseline measures: obtained by selecting the article's title as GeneRIF.

Because of these poor results, we attempt to combine both argumentative and title features together. In addition, we also try to shorten the candidate GeneRIF by removing a few words at the beginning or at the end of the candidate GeneRIF segment.

Features fusion

In this second approach, we decide to use the title as default GeneRIF. Then, we compute a Dice-like distance between candidates GeneRIF (which were classified as PURPOSE or CONCLUSION by the LASt system) and titles, as provided in MedLine citations. Again, in this approach, GPE tokens are overweighted in the Dice calculus.

Run	Sent.	Dice1	Dice2	Dice3	Dice4
1	14	52.78	54.33	37.72	40.65
2	23	52.41	54.22	37.61	40.44
3	31	51.06	52.84	35.43	38.70
Without using the string length of candidates GeneRIFs					
4	14	51.98	53.91	36.82	39.60

Table 6. Final results: combining features from the title and argumentative features for 3 different thresholds (run 1 to 3). We also observe that shortening strategies results in a modest improvement (run 1 vs. 4).

For GPE identification, we extracted a list of synonyms from LocusLink for each of the targeted Locus: thus, for the query 11 (Locus ID =7066), the list of synonyms (or related terms) is the following: *THPO*, *thrombopoietin (myeloproliferative leukemia virus oncogene ligand, megakaryocyte growth and development factor), TPO, MGDF, MKCSF, MPLLG*. We ran the system with different threshold. Varying this threshold results in changing the proportion of candidate GeneRIF extracted from the abstract vs. the article's title. In table 6, three of the most interesting results are reported. The top performing run, run 1, is the one we submitted. Although very empirical and so data-driven, these thresholds were found particularly stable, and calculating the threshold for run 1 on half of the data did not result in any degradation of performances for the information extraction task.

Reducing GeneRIFs' Length

The sentence compression step can be seen as a word removal process: it combines syntactic features (based on an hybrid part-of-speech tagger [8]), a set of ad hoc triggers (such as "In this paper…") and argumentative structuring. The basic syntactic removal attempts to remove non-content bearing clauses: phrases introducing relative clauses ("data suggest that", "these data indicate that", "in this report, we provide the first direct evidence that"…) and other introduction/adverbial clauses ("in addition", "surprisingly", "finally"…) are thus removed. Finally, clauses (from 3 to 7 words) expressing methods are also filtered out when found at the extremities of the candidate segment: for example, phrases such as "…using this method…" are removed. Serializing these compression strategies results in removing clauses as long as "These data indicate that, in contrast to current models…". It is to be noted that clauses containing GPE do not follow the length reduction process because such segments are potentially relevant: for instance, the segment "ARH is a modular adaptor protein that…" is not removed, because ARH is identified as a GPE.

Conclusion

In conclusion, our preliminary observations suggest that structural features (those stemming from the explicit structure of MedLine citations, such as the title, as well as those extracted from the latent structure, such as the argumentative structure) must be seen as a reasonable step in direction of automatic GeneRIF extraction. However, the task will require additional materials, as well as more powerful fusion's strategies, as explored in Savoy and Perret [15]. It is to be noted that for the secondary task, it was not clear whether other GeneRIF were to be used as training instances or not; however it is to be noted that apart from our experiments, other competitive approaches [15] [16] [17] were based on classifiers trained on GeneRIF. The use of the "function" axe in the Gene Ontology [7] together with using better gene and protein names recognition tools [21] [22] could also help identifying gene functions in MedLine abstracts.

Finally, we would like to remark that the chosen metrics were sufficient to compare the different approaches, but that more elaborated -and unfortunately more human-intensive- measures should be investigated in order to take into account the lexical variation of the biomedical language in general and of gene and protein names in particular [18].

Acknowledgments

The study reported in this paper has been partially supported by the Swiss National Foundation within the IM2 NCCR (http://www.im2.ch/) and by a separate SNF grant (MedTag, Grant 3200-065228.01). We used the CMU Statistical

[4] Best = 57.83; Median = 49.31, among all submitted runs and systems for the TREC genomic track (task II).

Language Modeling toolkit to compute the n-grams frequency of the argumentative classifier. We would like to thank Frédérique Lisacek, Arnaud Gaudinat, Johann Marty and Thomas de la Charrière for the comments they provided on this study.

References

[1] P Ruch, R Baud, and A Geissbühler. Learning-free Text Categorization, In *AIME* 2003, Springer LNCS/LNAI 2780, Dojat, M; Keravnou, E; Barahona, P (Eds.).

[2] 1a A Aronson, O Bodenreider, H Chang, S Humphrey, J Mork, S Nelson, T Rindflesch, W Wilbur, The NLM Indexing Initiative. Proc AMIA Symp. 2000. p. 17-21.

[3] H Edmundson, New methods in automatic abstracting. Journal of The Association for Computing Machinery, 16(2), pp. 264-285. Reprinted in Mani, I., and Maybury, M., eds., Advances in Automatic Text Summarization, MIT Press, p. 21-42. 1969.

[4] C Orasan, Patterns in scientific abstracts, in Proceedings of Corpus Linguistics 2001 Conference, Lancaster University, Lancaster, UK, pp. 433 - 443, 2001.

[5] J Swales, Genre Analysis: English in academic and research settings, Cambridge University Press, 1990.

[6] F Salanger-Meyer, Discoursal movements in medical English abstracts and their linguistic exponents: a genre analysis study, INTERFACE: Journal of Applied Linguistics 4(2), 1990, pp. 107 - 124

[7] F Ehrler and P Ruch. Report on the BioCreative Experiment: Task Presentation, System Description and Preliminary Results. In *Notebook of BioCreative 2004*, Granada, March 2004. (to appear)

[8] P Ruch, R Baud, P Bouillon, and G Robert. Minimal Commitment and Full Lexical Disambiguation: Balancing Rules and Hidden Markov Models. *CoNLL-2000 (ACL-SIGNLL) Proceedings*, p. 111-115, 2000.

[9] J Kupiec, J Pedersen and F Chen: A Trainable Document Summarizer, SIGIR 1995, p. 55-60.

[10] S Teufel and M Moens, Argumentative classification of extracted sentences as a first step towards flexible abstracting. In: I. Mani, M. Maybury (eds.), Advances in automatic text summarization, MIT Press, 1999.

[11] S Teufel and M Moens, Sentence Extraction and rhetorical classification for flexible abstracts, AAAI Spring Symposium on Intelligent Text summarization, 1998, p. 89-97.

[12] C Nedellec, M Ould Abdel Vetah, Ph Bessières: Sentence Filtering for Information Extraction in Genomics, a Classification Problem. PKDD 2001: 326-337

[13] B de Bruijn and J Martin, Finding Gene Function using LitMiner. In *Notebook of the TREC-2003*, p. 486-494.

[14] M Kayaalp, AR Aronson, SM Humphrey, N ide, L Tanabe, LH Smith, R Looane, JG Mork and O Bodenreider. Methods for Accurate Retrieval of MedLine Citations in Functional Genomics. In *Notebook of the TREC-2003*, p. 175-184.

[15] J Savoy, Y Rasolofo and L Perret. Report on the TREC 2003 Experiment: Genomics and Web Searches. In *Notebook of the TREC-2003*, Gaithersburg, p. 686-697.

[16] G Bhalotia, P Nakov, A Schwartz and M Hearst. Biotext Team Report for the TREC2003 Genomics Track. In *Notebook of the TREC-2003*, p. 158-166.

[17] R Jeller, M Schuemie, C vander Eijk, M Weeber, E van Mulligen, B Schijvenaars, B Mons and K Kors. Searching for GeneGIFs: Concept-based Query Expansion and Bayesian Cassification. In *Notebook of the TREC-2003*, 167-174.

[18] S Albert, S Gaudan, H Knigge, A Raetsch, A Delgado, B Huhse, H Kirsch, M Albers, D Rebholz-Schuhmann and M Koegl. Computer-assisted generation of a protein-interaction database for nuclear receptors. *Mol Endocrinol*. 2003 Aug;17(8):1555-67.

[19] W Hersh and RT Bhupatiraju. TREC genomics track overview. In *Notebook of the TREC-2003*, 148-157.

[20] Mitchell, J.A., Aronson, A.R., Mork, J.G., Folk, L.C., Humphrey, S.M., and Ward, J.M. Gene indexing: characterization and analysis of NLM's GeneRIFs. In *Proceedings of the AMIA 2003 Annual Symposium. 2003*.

[21] Collier, N., Nobata, C., and Tsujii, J. "Extracting the Names of Genes and Gene Products with a Hidden Markov Model", In COLING, pp. 201-207, August, 2000.

[22] Blaschke, C., Andrade, M.A., Ouzounis, C.A., and Valencia, A. Automatic extraction of biological information from scientific text: protein-protein interactions. In *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology*, 1999, p. 60-67.

[23] S Blott, C Gurrin, GJF Jones, AF Smeaton and T Sodring. On the Use of MeSH Headings to Improve Retrieval Effectiveness. In *Notebook of the TREC-2003*, Gaithersburg, p. 355-364.

[24] PB Dobrokhotov, C Goutte, AL Veuthey, E Gaussier. A Probabilistic Information Retrieval Approach to Medical Annotation in SWISS-PROT. *Stud Health Technol Inform*.2003;95:421-6.

CONCLUSION

CONCLUSION|00155670|data suggest that the lack of an LXR element in the region from -56 to -49 of the human CYP7A1 promoter may account for some of the differences in response to diets between humans and rodents|PURPOSE=00161186|METHODS=00164845|RESULTS=00164080|CONCLUSION=00155670|

CONCLUSION|00159040|conclude that E2F proteins and Sp1 play an important role in the control of p18 expression|PURPOSE=00163088|METHODS=00168915|RESULTS=00167241|CONCLUSION=00159040|

CONCLUSION|00159486|The anti-apoptotic activity of IL-4 in B cells is mediated through the activation of Stat6 and subsequent transcription of Bcl-xL.|PURPOSE=00160406|METHODS=00162612|RESULTS=00163409|CONCLUSION=00159486|

CONCLUSION|00160383|Data suggest that increased activity of mutated interleukin 3 is due to a change from a rare ligand to a common one, allowing the increase in IL-3-dependent signaling.|PURPOSE=00163992|METHODS=00167497|RESULTS=00166006|CONCLUSION=00160383|

CONCLUSION|00161952|SHD1 of Slac2-a/melanophilin alone is both necessary and sufficient for high affinity specific recognition of the GTP-bound form of Rab27A|PURPOSE=00163587|METHODS=00166343|RESULTS=00165666|CONCLUSION=00161952|

CONCLUSION|00162239|The C-terminus of Slac2-a/melanophilin contains a novel actin-binding site, which may be involved in capture of Rab27-containing organelles in the actin-enriched cell periphery.|PURPOSE=00164858|METHODS=00168703|RESULTS=00168628|CONCLUSION=00162239|

CONCLUSION|00162489|data suggest that TTF-1 plays an important regulatory role in the gene transcription for pituitary adenylate cyclase-activating polypeptide|PURPOSE=00165883|METHODS=00171795|RESULTS=00170670|CONCLUSION=00162489|

CONCLUSION|00162908|method for identifying both the alpha- and beta-chains of the T cell receptor (TCR) from individual pancreatic islet-infiltrating T cells at the earliest stages of disease in nonobese diabetic mice (NOD)|PURPOSE=00163946|METHODS=00165852|RESULTS=00166164|CONCLUSION=00162908|

CONCLUSION|00163248|results indicate that the GnRH receptor activates both G(q) and G(s) signaling to regulate gene expression in L beta T2 cells|PURPOSE=00166755|METHODS=00167836|RESULTS=00166986|CONCLUSION=00163248|

CONCLUSION|00163599|In the case of Fas-mediated apoptosis, when we transiently introduced these hybrid-ribozyme libraries into Fas-expressing HeLa cells, we were able to isolate surviving clones that were resistant to or exhibited a delay in Fas-mediated apoptosis|PURPOSE=00165907|METHODS=00166384|RESULTS=00164822|CONCLUSION=00163599|

CONCLUSION|00163615|results suggest that Wengen can act as a component of a functional receptor for Eiger|PURPOSE=00165475|METHODS=00169040|RESULTS=00169159|CONCLUSION=00163615|

CONCLUSION|00163747|Sp1 plays a role in regulation of promoter activity and in PKA-mediated expression of mitochondrial serine:pyruvate aminotransferase|PURPOSE=00165356|METHODS=00169810|RESULTS=00169495|CONCLUSION=00163747|

CONCLUSION|00163764|findings suggest that PTP1B modulates insulin signaling in liver and fat, and that therapeutic modalities targeting PTP1B inhibition may have clinical benefit in type 2 diabetes|PURPOSE=00165438|METHODS=00169999|RESULTS=00169256|CONCLUSION=00163764|

CONCLUSION|00163920|DIAP1 is required to prevent excess accumulation of the first form of processed DRONC, presumably through its ability to act as a ubiquitin-protein ligase|PURPOSE=00166270|METHODS=00168954|RESULTS=00169371|CONCLUSION=00163920|

CONCLUSION|00163964|redundancy in the functions of PPARs alpha and delta as transcriptional regulators of fatty acid homeostasis and suggest that in skeletal muscle high levels of the delta-subtype can compensate for deficiency of PPAR alpha|PURPOSE=00165519|METHODS=00167739|RESULTS=00166944|CONCLUSION=00163964|

CONCLUSION|00164059|Results suggest that Bcl-2 activates NF-kappa B by a signaling mechanism that involves Raf-1/MEKK-1 mediated activation of IKK beta.|PURPOSE=00167731|METHODS=00170060|RESULTS=00170330|CONCLUSION=00164059|

CONCLUSION|00164077|a short sequence present in the N-terminal domain has a role in controlling anterograde trafficking of ionotropic glutamate receptors|PURPOSE=00164949|METHODS=00168753|RESULTS=00168016|CONCLUSION=00164077|

PURPOSES

PURPOSE|00160209|inactivation sensitizes cells to apoptosis via an increase of both p14ARF and p53 levels and an alteration of the Bax/Bcl-2 ratio|PURPOSE=00160209|METHODS=00162198|RESULTS=00160962|CONCLUSION=00160414|

PURPOSE|00160467|The structure of human mini-TyrRS containing both the catalytic & the anticodon recognition domains, is reported to a resolution of 1.18 A. The spatial disposition of the anticodon recognition domain relative to the catalytic domain is unique.|PURPOSE=00160467|METHODS=00162177|RESULTS=00162948|CONCLUSION=00160494|

PURPOSE|00161526|demonstrates role of the Sp1 protein in basal and estrogen-induced growth and gene expression in breast cancer|PURPOSE=00161526|METHODS=00166143|RESULTS=00165948|CONCLUSION=00163118|

PURPOSE|00163162|Reentrant loop II of the GLT-1 transporter forms part of an aqueous pore, the access of which is blocked by the glutamate analogue dihydrokainate, and that sodium influences the conformation of this pore-loop.|PURPOSE=00163162|METHODS=00165455|RESULTS=00165612|CONCLUSION=00163739|

PURPOSE|00164580|restoring FoxM1B expression in old-aged mice caused elevated levels of Cyclin B1, Cyclin B2, Cdc25B, Cdk1, and p55CDC mRNA as well as stimulating Cdc25B nuclear localization during liver regeneration, all of which are required for mitosis|PURPOSE=00164580|METHODS=00164982|RESULTS=00165057|CONCLUSION=00164983|

PURPOSE|00164784|role in activating the JNK and p38 MAP kinase cascades in response to environmental stresses such as reactive oxygen species|PURPOSE=00164784|METHODS=00169071|RESULTS=00168713|CONCLUSION=00164983|

PURPOSE|00164823|role in regulating transcription of the matrix metalloproteinase-9 gene induced by IL-1 and TNF-alpha in glioma cells via NF-kappa B|PURPOSE=00164823|METHODS=00167145|RESULTS=00167318|CONCLUSION=00165231|

PURPOSE|00165894|promotes survival of lung cancer cells by suppressing apoptosis through dysregulation of the mitochondrial caspase pathway|PURPOSE=00165894|METHODS=00168480|RESULTS=00168750|CONCLUSION=00167276|

PURPOSE|00166102|Ets-1 and Sp1 have a role in regulating FasL expression in human vascular smooth muscle cells|PURPOSE=00166102|METHODS=00171102|RESULTS=00169973|CONCLUSION=00166540|

PURPOSE|00166623|Inactivation of p21WAF1 sensitizes cells to apoptosis via an increase of both p14ARF and p53 levels and an alteration of this protein and Bcl-2 ratio|PURPOSE=00166623|METHODS=00168861|RESULTS=00167661|CONCLUSION=00167054|

PURPOSE|00166976|keratinocyte growth factor (KGF), a key stimulator of epithelial cell proliferation during wound healing, preferentially binds to collagens I, III, and VI.|PURPOSE=00166976|METHODS=00167817|RESULTS=00168188|CONCLUSION=00167682|

PURPOSE|00167254|Nrdp1/FLRF is a ubiquitin ligase promoting ubiquitination and degradation of this epidermal growth factor receptor family member|PURPOSE=00167254|METHODS=00171250|RESULTS=00172377|CONCLUSION=00168888|

METHODS

METHODS|00169100|Foxm1b transcription factor regulates expression of cell cycle proteins essential for hepatocyte entry into DNA replication and mitosis.|PURPOSE=00169438|METHODS=00169100|RESULTS=00170338|CONCLUSION=00169513|

METHODS|00171985|Cleavage of p21waf1 by proteinase-3, a myeloid-specific serine protease, potentiates cell proliferation|PURPOSE=00172082|METHODS=00171985|RESULTS=00172172|CONCLUSION=00172101|

METHODS|00173131|activation by furin via one of two consecutive recognition sites|PURPOSE=00174592|METHODS=00173131|RESULTS=00174404|CONCLUSION=00174936|

RESULTS

RESULTS|00162783|apoE binds to the LDL receptor by interacting with more than one of the receptor ligand-binding repeats.|PURPOSE=00163421|METHODS=00164015|RESULTS=00162783|CONCLUSION=00162798|

RESULTS|00171020|signals to mitochondria via FADD, caspase-8/10, Bid, and Bax but differentially regulate events downstream from truncated Bid compared to TRAIL receptor 2|PURPOSE=00171860|METHODS=00171183|RESULTS=00171020|CONCLUSION=00171981|

Annex 1. Samples of GeneRIFs after argumentative classification: the first row gives the class; the second row gives the score of the selected class; then the textual segments is given; finally the score of the other classes is indicated.