Software

Deafness mutation mining using regular expression based pattern matching

Christopher M Frenz

Address: Department of Computer Engineering Technology, New York City College of Technology (CUNY), 300 Jay St, Brooklyn, NY 11201, USA

Email: Christopher M Frenz - cfrenz@gmail.com

Abstract

Background: While keyword based queries of databases such as Pubmed are frequently of great utility, the ability to use regular expressions in place of a keyword can often improve the results output by such databases. Regular expressions can allow for the identification of element types that cannot be readily specified by a single keyword and can allow for different words with similar character sequences to be distinguished.

Results: A Perl based utility was developed to allow the use of regular expressions in Pubmed searches, thereby improving the accuracy of the searches.

Conclusion: This utility was then utilized to create a comprehensive listing of all DFN deafness mutations discussed in Pubmed records containing the keywords "human ear".

Background

Biological research has yielded a vast amount of research data, which can often provide novel insights when the data can be viewed in an aggregated fashion, and thus recent studies have employed computational methods of information extraction from the biomedical literature. These studies have dealt with a wide range of information extractions, including the names of genes and proteins[1], intermolecular relationships [2], and molecular biological descriptors [3].

Pubmed currently catalogs citation and abstract information for over 4,400 biomedical research journals and houses a citation database of over 12.8 million citations[4]. With any database of this size the return of relevant query results is often a difficult task, given the large number of potential matches there likely are for any single query term. These difficulties are compounded even further, given that Pubmed records are all natural language records and searches cannot readily be conducted using a predefined set of terms, as is the case for many relational databases. Thus Pubmed employs a word-matching algorithm, which seeks to match query words to the contents of citation records, and will return all records containing that word in their order of publication starting with the most recent.

For certain types of queries, such as mutations, basic word matching is an ineffective search strategy, since an effective query cannot be specified as a single word, but rather is better expressed as a textual pattern, such as [Residue] [Position] [MutantResidue] [5]. The use of textual pattern matching, however, has a wide array of uses that extend beyond the location of mutations within Pubmed records, and include the ability to distinguish between articles which discuss pKa values as opposed to articles that discuss Protein Kinase A (PKA), which would both be yielded by a Pubmed search for the "pKa" word. These above examples, illustrate the two major applications that text patterns offer to Pubmed searching; 1) the identification

of elements that cannot be specified by a single word and 2) distinguishing between two different words that are comprised of a similar sequence of characters [6]. Textual patterns are commonly matched via the use of regular expressions and studies that involve the extraction of biochemical mutation data from biomedical literature have demonstrated a high degree of success [5,7]. This study seeks to develop a Perl based utility Perl Regular Expressions for Pubmed (PREP.pl, See Additional File 1, which allows the searching of Pubmed citation records for the presence of textual patterns and for the placement of match containing records into an HTML formatted output file. This Perl based utility will then be utilized to construct a comprehensive listing of DFN mutations discussed in Pubmed records containing the "human ear" keywords.

Implementation
The PREP utility
The script interacts with Pubmed via NCBI's E-Utilities interface [4] and the LWP module handles all HTTP based communication. The script begins by using the ESearch method to query Pubmed for all records containing a user defined search term, such as "lysozyme" or "HIV". Pubmed ID numbers of all matching records are temporarily stored on the Pubmed server and can be accessed using the EFetch method and an assigned Web environment variable and query key, which is returned by the ESearch method. Records returned by the EFetch method are requested in XML format, since the well-defined hierarchical structures of XML documents greatly simplifies parsing tasks [6]. This script makes use of the XML::LibXML Perl module for XML parsing, and from each Pubmed record the title of the article, the journal information, the abstract, and the Pubmed ID of the record, are extracted, based on their corresponding XML tag names.

A user specified regular expression is then used to search the abstract and title fields of each record and look for a textual pattern match. Only the title and abstract fields are searched, since these are the fields in which pattern matches are most likely to be found, and the elimination of other fields reduces the potential for false positives. If a match occurs the journal information, the abstract, and title are output to an HTML file (Figure 1).

The title is output in the format of a hyperlink to the Pubmed record that corresponds to that article, to allow for easy retrieval of any additional information pertaining to the article that the output file does not provide or in certain cases easy retrieval of the entire article. The generation of an HTML output allows for the results to be easily shared among users working on disparate computing platforms. Records that contain no matches to the text pattern

of interest are not written to an output file. On an AMD Athlon 2000+ the PREP script can process an average of 500 abstracts per minute.

PREP can be run from the command line of any Linux or Unix machine that has the XML::Lib::XML Perl module installed. The regular expression used within the script is modified by changing the value of the $regex variable within the script, as indicated by the code documentation. Command line script execution can be initiated using the standard Perl command line syntax of "perl Prep.pl Keywords". The utility was chosen to be implemented in a command line fashion since this makes the utility suitable for easy inclusion in more comprehensive data mining scripts where the search functionality of PREP may provide useful.

Validation of utility
As a test of the specificity obtainable by the PREP script, all Pubmed records that resulted from a search for the word "lysozyme" were checked for pattern matches to the Protein Kinase A abbreviation "PKA" by using the regular expression PKA within the PREP script. At the time the test was conducted there were 19,964 records returned, and the PREP script indicated that only 3 records contained the textual pattern "PKA". These findings were manually confirmed by going through all records returned by the lysozyme search. The textual pattern "pKa", however, is actually fairly common throughout the lysozyme record set, and the PREP script successfully eliminated these "pKa" containing records from the search results, whereas the standard Pubmed keyword search is unable to accomplish this. Thus, this test is indicative that with a well-formed regular expression a high degree of specificity and search refinement can be achieved between different words with like character compositions. While no false positives were noted during the manual confirmation of these search results, the potential source of false negatives for this search would be abstracts that discussed Protein Kinase A without mentioning the abbreviation PKA.

The ability of the PREP script to identify elements that cannot be specified as a single word was tested by searching for mutations in the records returned by a search for "hen egg white lysozyme" using the regular expression:

[ARNDCEQGHILKMFPSTWYV]\d+[ARNDCEQGHILKM-FPSTWYV]|

[A-Z][a-z][a-z]\d+[A-Z][a-z][a-z]

This expression allows for the identification of mutations written out in both the single letter amino acid notation as well as the three-letter notation. The "hen egg white lysozyme" search of Pubmed yielded 1146 records of which
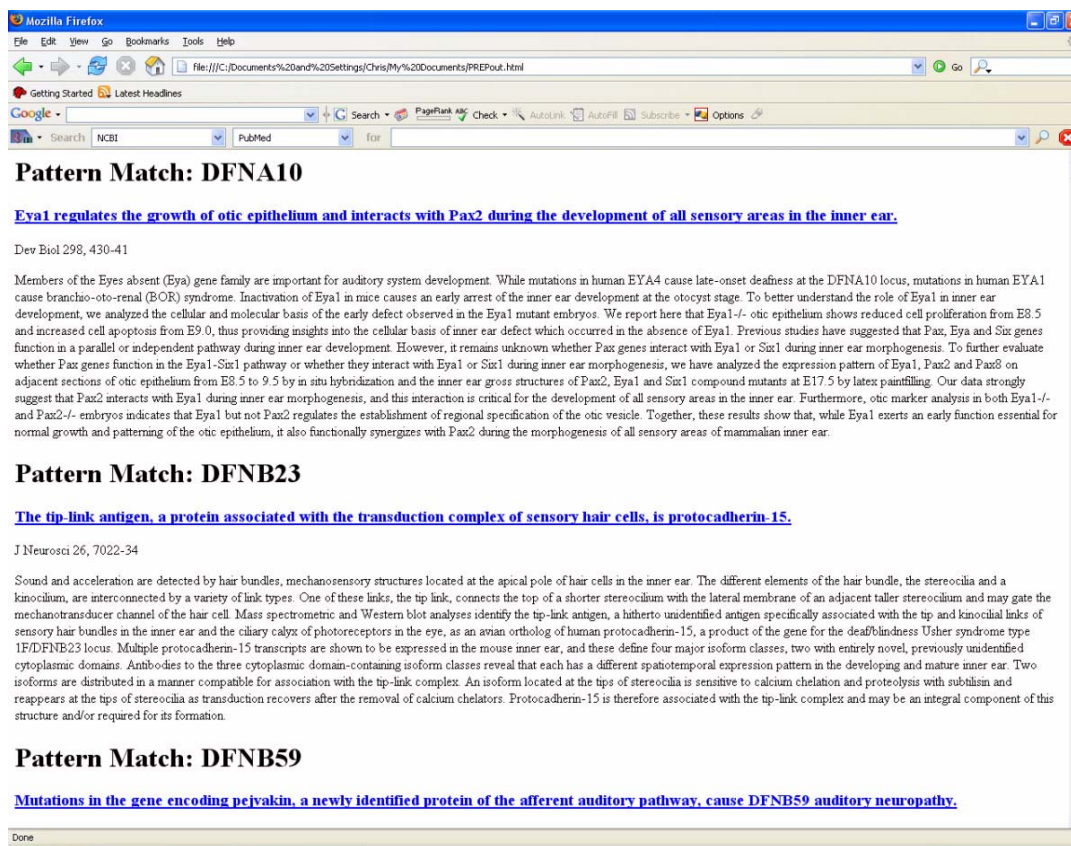
Figure 1
A screen shot of the PREP program output for a search for DFN deafness.

PREP identified 62 as matching the above regular expression pattern, and were manually confirmed. Of these 62 matches, 36 (58%) records contained actual mutations while the remaining 42% contained false positives, such as the abbreviation for T4 Lysozyme (T4L). In order to lessen the percentage of false positives, the false positives were examined and it became apparent that many of the same false positives occurred in repeated records. Thus a simple filter was created by defining a second regular expression, which explicitly matched the false positives, and prevented them from being recorded in the program output, thereby eliminating these repeating false positives. In this manner, the total number of PREP matches was reduced to 47, raising the percentage of valid positives up to 77% and reducing the number of false negatives to 23%. This is indicative that the PREP script can be an effective tool in reducing the search space necessary for manual processing by taking the 1146 initial records and narrowing down the list of possible records to 47, or 4% of the original search space. It should be further noted that the PREP script did not miss any records that contained matching patterns within the data set, and that the DFN prefix associated with deafness mutations is less likely to turn up false positives than the more generalized pattern associated with biochemical mutation data. This validation exercise, though, does demonstrate the utility of an application specific filter as a means of reducing false positives where warranted.

Results
The textual pattern DFN [A-Z]\d+ was defined, where [A-Z] could be any letter between A and Z and \d+ could be a combination of one or more numeric digits and this pattern used to search through records returned by a PubMed search for the keywords "human ear". The search yielded 61,371 Pubmed records and out of those 117 contained a pattern match. All of the pattern matches corresponded to valid DFN deafness mutations and no false positives were returned. The DFN mutation found in the 117 matching records are summarized in Table 1. In cases where multiple records discussed a mutation, a representative record is listed in the source field, rather than every record, to limit table length.

Discussion & conclusion

The PREP script was able to process 61,371 Pubmed records displaying the keywords "human ear" and narrow the relevant search space down to 117 articles that contain different DFN deafness mutations. This is slightly less than 0.2% of the original search space, demonstrating the utility of pattern matching in aiding researchers in obtaining relevant information from the biomedical literature. Furthermore, the lack of false positives among the returned results demonstrates that the accuracy and utility of this approach can be further enhanced when the

defined pattern possesses a high degree of specificity. The PREP approach to literature searching would therefore allow researchers to uncover a diversity of information pertaining to deafness mutations in a single search, whereas uncovering the same 45 DFN deafness mutations (Table 1) by standard keyword searches would take considerably more time and effort. However, when utilizing such an approach to literature searching, in addition to false positives, it is important to carefully consider the keywords presented to Pubmed. For example, this "human ear" keyword search failed to uncover the

Table 1: Mutations located in the PREP program results

Mutation Name	Mutation Effect	Representative Source
DFNA1	Diaphonous gene mutation associated with autosomal dominant non-syndromic hearing loss	9
DFNA10	Mutation in EYA4 causes late onset deafness	10
DFNA11	MYO7A mutation that results in progressive loss of mechanotransduction	11
DFNA12	TECTA mutation resulting in hearing impairment	12
DFNA13	Mutation leading to cochlear conductive loss	13
DFNA14	wolframin mutation cauisng non-syndromic dominant low frequency hearing loss	14
DFNA15	Mutation in POU4F3 that leads to autosomal dominant non-syndromic hearing loss	15
DFNA17	Mutation in myosin heavy chain IX linked to hearing impairment	16
DFNA2	KCNQ4 potassium channel mutation leading to progressive hearing loss	17
DFNA20	ACTG1 mutation causing autosomal dominant heairng loss	18
DFNA24	Caspase 3 mutation associated with autosomal dominant non-syndromic hearing loss	19
DFNA26	ACTG1 mutation causing autosomal dominant heairng loss	18
DFNA36	Mutation in transmembrane cochlear expressed gene 1 causing progressive deafness	20
DFNA38	wolframin mutation cauisng non-syndromic dominant low frequency hearing loss	14
DFNA39	Hearing loss associated with Dentinogenesis imperfecta	21
DFNA4	MYH14 mutation leading to autosomal dominant hearing loss	23
DFNA48	MYO1A mutation resulting in autosomal dominant hearing loss	22
DFNA5	Mutation causing autosomal dominant hearing impairment	24
DFNA6	wolframin mutation cauisng non-syndromic dominant low frequency hearing loss	14
DFNA8	TECTA mutation resulting in hearing impairment	12
DFNA9	Coagulation factor C homology gene mutations causing sensioneural hearing loss	25
DFNB1	Connexin 26 mutation leading to non-syndromic hearing loss	26
DFNB11	Mutation in transmembrane cochlear expressed gene 1 causing congenital deafness	20
DFNB12	Cadherin 23 mutation causing prelingual hearing loss	14
DFNB13	Mutations causing autosomal recessive non-syndromic deafness	27
DFNB14	Hearing loss associated with split hand/split foot malformation	28
DFNB16	Stereocilan mutation leading to autsomal recessive non-syndromic deafness	29
DFNB17	FAM3C mutation causing autosomal recessive non-syndromic hearing loss	30
DFNB18	Deafness associated with Usher syndrome 1C	31
DFNB2	Deafness associated with mutations in myosin VIIA gene	32
DFNB22	Otoancorin mutation resulting in autosmal recessive deafness	33
DFNB23	Usher Syndrome 1F related deafness	34
DFNB25	Chromosome 5 mutation that effects sensory mechanotransduction	35
DFNB28	TRIOBP mutation resulting in recessive prelingual sensioneural hearing loss	36
DFNB29	CLDN14 mutations resulting in autosomal recessive non-syndromic deafness	37
DFNB3	Myo15a related non-syndromic deafness	38
DFNB30	Mutation in myosin IIIA resulting in progressive hearing loss	39
DFNB31	Whirlin mutation resulting in hearing loss	18
DFNB4	Mutation in PDS gene causing congenital deafness	40
DFNB59	Autosomal recessive auditory neuropathy	41
DFNB6	Mutation causing autosomal recessive deafness	42
DFNB67	Mutation in THMS causing recessive non-syndromic hearing loss	43
DFNB7	Mutation in transmembrane cochlear expressed gene 1 causing congenital deafness	20
DFNB8	TMPRSS3 mutation associated with non-syndromic autosommal recessive hearing loss	44
DFNB9	Mutation in otoferlin causing prelingual hearing loss	45

DFNB35 mutation [8] since it does not appear in an abstract that contains the words "human" and "ear". Potential sources of false negatives among search results include papers that do not utilize the DFN based nomenclature to discuss the mutation or articles that mention the abbreviation in the text, but not the abstract. Based on the validation tests, however, the false negative rate is expected to be low. Even with these limitations, however, the textual pattern based search methodology presented here can be of great value to researchers in the otolaryngological sciences as well as in other biomedical disciplines, since regular expressions can also be created to match other biological patterns, such as DNA or protein sequences, ions, enzyme names, and numerous other possibilities.

Availability & requirements
Project Name: PREP: Perl Regular Expressions for PubMed

Project Home Page: http://bioinformatics.org/project/?group_id=494

Operating Systems: Linux/Unix

Programming Language: Perl

Other Requirements: XML::LibXML Perl Module

License: Perl Artistic License

Competing interests
The author(s) declare that they have no competing interests.

Authors' contributions
CMF is responsible for the study and manuscript in their entirety.

Additional material

Additional file 1
The PREP.pl Perl script.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1472-6947-7-32-S1.pl]

Acknowledgements

References
1. Leonard JE, Colombe JB, Levy JL: **Finding relevant references to genes and proteins in Medline using a Bayesian approach.** *Bioinformatics* 2002, **18:**1515-1522.
2. Yoshida M, Fukuda K, Takagi T: **PNAD-CSS: a workbench for constructing a protein name abbreviation dictionary.** *Bioinformatics* 2000, **16:**169-175.
3. Andrade MA, Bork P: **Automated extraction of information in molecular biology.** *FEBS Lett* 2000, **476:**12-17.
4. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Helmberg W, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pontius JU, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2005, **33:**D39-D45.
5. Horn F, Lau AL, Cohen FE: **Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors.** *Bioinformatics* 2004, **20:**557-568.
6. Frenz CM: *Pro Perl Parsing* New York: Springer-Verlag; 2005.
7. Rebholz-Schuhmann D, Marcel S, Albert S, Tolle R, Casari G, Kirsch H: **Automatic extraction of mutations from Medline and cross-validation with OMIM.** *Nucleic Acids Res* 2004, **32:**135-142.
8. Ansar M, Din MA, Arshad M, Sohail M, Faiyaz-Ul-Haque M, Haque S, Ahmad W, Leal SM: **A novel autosomal recessive non-syndromic deafness locus (DFNB35) maps to 14q24.1–14q24.3 in large consanguineous kindred from Pakistan.** *Eur J Hum Genet* 2003, **11:**77-80.
9. Lalwani AK, Jackler RK, Sweetow RW, Lynch ED, Raventos H, Morrow J, King MC, Leon PE: **Further characterization of the DFNA1 audiovestibular phenotype.** *Arch Otolaryngol Head Neck Surg* 1998, **124:**699-702.
10. Zou D, Silvius D, Rodrigo-Blomqvist S, Enerback S, Xu PX: **Eya1 regulates the growth of otic epithelium and interacts with Pax2 during the development of all sensory areas in the inner ear.** *Dev Biol* 2006, **298:**430-441.
11. Bolz H, Bolz SS, Schade G, Kothe C, Mohrmann G, Hess M, Gal A: **Impaired calmodulin binding of myosin-7A causes autosomal dominant hearing loss (DFNA11).** *Hum Mutat* 2004, **24:**274-275.
12. Verhoeven K, Van Laer L, Kirschhofer K, Legan PK, Hughes DC, Schatteman I, Verstreken M, Van Hauwe P, Coucke P, Chen A, Smith RJ, Somers T, Offeciers FE, Van de Heyning P, Richardson GP, Wachtler F, Kimberling WJ, Willems PJ, Govaerts PJ, Van Camp G: **Mutations in the human alpha-tectorin gene cause autosomal dominant non-syndromic hearing impairment.** *Nat Genet* 1998, **19:**60-62.
13. De Leenheer EM, Bosman AJ, Kunst HP, Huygen PL, Cremers CW: **Audiological characteristics of some affected members of a Dutch DFNA13/COL11A2 family.** *Ann Otol Rhinol Laryngol* 2004, **113:**922-929.
14. McHugh RK, Friedman RA: **Genetics of hearing loss: Allelism and modifier genes produce a phenotypic continuum.** *Anat Rec A Discov Mol Cell Evol Biol* 2006, **288:**370-381.
15. Hertzano R, Montcouquiol M, Rashi-Elkeles S, Elkon R, Yucel R, Frankel WN, Rechavi G, Moroy T, Friedman TB, Kelley MW, Avraham KB: **Transcription profiling of inner ears from Pou4f3(ddl/ddl) identifies Gfi1 as a target of the Pou4f3 deafness gene.** *Hum Mol Genet* 2004, **13:**2143-2153.
16. Parker LL, Gao J, Zuo J: **Absence of hearing loss in a mouse model for DFNA17 and MYH9-related disease: the use of public gene-targeted ES cell resources.** *Brain Res* 2006, **1091:**235-242.
17. Kharkovets T, Dedek K, Maier H, Schweizer M, Khimich D, Nouvian R, Vardanyan V, Leuwer R, Moser T, Jentsch TJ: **Mice with altered KCNQ4 K+ channels implicate sensory outer hair cells in human progressive deafness.** *EMBO J* 2006, **25:**642-652.
18. van Wijk E, Krieger E, Kemperman MH, De Leenheer EM, Huygen PL, Cremers CW, Cremers FP, Kremer H: **A mutation in the gamma actin 1 (ACTG1) gene causes autosomal dominant hearing loss (DFNA20/26).** *J Med Genet* 2003, **40:**879-884.
19. Morishita H, Makishima T, Kaneko C, Lee YS, Segil N, Takahashi K, Kuraoka A, Nakagawa T, Nabekura J, Nakayama K, Nakayama KI: **Deafness due to degeneration of cochlear neurons in cas-**

pase-3-deficient mice. *Biochem Biophys Res Commun* 2001, **284:**142-149.

20. Marcotti W, Erven A, Johnson SL, Steel KP, Kros CJ: **Tmc1 is necessary for normal functional maturation and survival of inner and outer hair cells in the mouse cochlea.** *J Physiol* 2006, **574:**677-698.

21. Xiao S, Yu C, Chou X, Yuan W, Wang Y, Bu L, Fu G, Qian M, Yang J, Shi Y, Hu L, Han B, Wang Z, Huang W, Liu J, Chen Z, Zhao G, Kong X: **Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP.** *Nat Genet* 2001, **27:**201-204.

22. Donaudy F, Ferrara A, Esposito L, Hertzano R, Ben-David O, Bell RE, Melchionda S, Zelante L, Avraham KB, Gasparini P: **Multiple mutations of MYO1A, a cochlear-expressed gene, in sensorineural hearing loss.** *Am J Hum Genet* 2003, **72:**1571-1577.

23. Donaudy F, Snoeckx R, Pfister M, Zenner HP, Blin N, Di Stazio M, Ferrara A, Lanzara C, Ficarella R, Declau F, Pusch CM, Nurnberg P, Melchionda S, Zelante L, Ballana E, Estivill X, Van Camp G, Gasparini P, Savoia A: **Nonmuscle myosin heavy-chain gene MYH14 is expressed in cochlea and mutated in patients affected by autosomal dominant hearing impairment (DFNA4).** *Am J Hum Genet* 2004, **74:**770-776.

24. Van Laer L, Pfister M, Thys S, Vrijens K, Mueller M, Umans L, Serneels L, Van Nassauw L, Kooy F, Smith RJ, Timmermans JP, Van Leuven F, Van Camp G: **Mice lacking Dfna5 show a diverging number of cochlear fourth row outer hair cells.** *Neurobiol Dis* 2005, **19:**386-399.

25. Robertson NG, Cremers CW, Huygen PL, Ikezono T, Krastins B, Kremer H, Kuo SF, Liberman MC, Merchant SN, Miller CE, Nadol JB, Sarracino DA, Verhagen WI, Morton CC: **Cochlin immunostaining of inner ear pathologic deposits and proteomic analysis in DFNA9 deafness and vestibular dysfunction.** *Hum Mol Genet* 2006, **15:**1071-1085.

26. Palmada M, Schmalisch K, Bohmer C, Schug N, Pfister M, Lang F, Blin N: **Loss of function mutations of the GJB2 gene detected in patients with DFNB1-associated hearing impairment.** *Neurobiol Dis* 2006, **22:**112-118.

27. Masmoudi S, Charfedine I, Rebeh IB, Rebai A, Tlili A, Ghorbel AM, Belguith H, Petit C, Drira M, Ayadi H: **Refined mapping of the autosomal recessive non-syndromic deafness locus DFNB13 using eight novel microsatellite markers.** *Clin Genet* 2004, **66:**358-364.

28. Fukushima K, Nagai K, Tsukada H, Sugata A, Sugata K, Kasai N, Kibayashi N, Maeda Y, Gunduz M, Nishizaki K: **Deletion mapping of split hand/split foot malformation with hearing impairment: a case report.** *Int J Pediatr Otorhinolaryngol* 2003, **67:**1127-1132.

29. Verpy E, Masmoudi S, Zwaenepoel I, Leibovici M, Hutchin TP, Del Castillo I, Nouaille S, Blanchard S, Laine S, Popot JL, Moreno F, Mueller RF, Petit C: **Mutations in a new gene encoding a protein of the hair bundle cause non-syndromic deafness at the DFNB16 locus.** *Nat Genet* 2001, **29:**345-349.

30. Pilipenko VV, Reece A, Choo DI, Greinwald JH: **Genomic organization and expression analysis of the murine Fam3c gene.** *Gene* 2004, **335:**159-168.

31. Johnson KR, Gagnon LH, Webb LS, Peters LL, Hawes NL, Chang B, Zheng QY: **Mouse models of USH1C and DFNB18: phenotypic and molecular analyses of two new spontaneous mutations of the Ush1c gene.** *Hum Mol Genet* 2003, **12:**3075-3086.

32. Ernest S, Rauch GJ, Haffter P, Geisler R, Petit C, Nicolson T: **Mariner is defective in myosin VIIA: a zebrafish model for human hereditary deafness.** *Hum Mol Genet* 2000, **9:**2189-2196.

33. Zwaenepoel I, Mustapha M, Leibovici M, Verpy E, Goodyear R, Liu XZ, Nouaille S, Nance WE, Kanaan M, Avraham KB, Tekaia F, Loiselet J, Lathrop M, Richardson G, Petit C: **Otoancorin, an inner ear protein restricted to the interface between the apical surface of sensory epithelia and their overlying acellular gels, is defective in autosomal recessive deafness DFNB22.** *Proc Natl Acad Sci USA* 2002, **99:**6240-6245.

34. Ahmed ZM, Goodyear R, Riazuddin S, Lagziel A, Legan PK, Behra M, Burgess SM, Lilley KS, Wilcox ER, Riazuddin S, Griffith AJ, Frolenkov GI, Belyantseva IA, Richardson GP, Friedman TB: **The tip-link antigen, a protein associated with the transduction complex of sensory hair cells, is protocadherin-15.** *J Neurosci* 2006, **26:**7022-7034.

35. Odeh H, Hagiwara N, Skynner M, Mitchem KL, Beyer LA, Allen ND, Brilliant MH, Lebart MC, Dolan DF, Raphael Y, Kohrman DC: **Char-**

36. **acterization of two transgene insertional mutations at pirouette, a mouse deafness locus.** *Audiol Neurootol* 2004, **9:**303-314.

36. Shahin H, Walsh T, Sobe T, Abu Sa'ed J, Abu Rayan A, Lynch ED, Lee MK, Avraham KB, King MC, Kanaan M: **Mutations in a novel isoform of TRIOBP that encodes a filamentous-actin binding protein are responsible for DFNB28 recessive nonsyndromic hearing loss.** *Am J Hum Genet* 2006, **78:**144-152.

37. Wilcox ER, Burton QL, Naz S, Riazuddin S, Smith TN, Ploplis B, Belyantseva I, Ben-Yosef T, Liburd NA, Morell RJ, Kachar B, Wu DK, Griffith AJ, Riazuddin S, Friedman TB: **Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29.** *Cell* 2001, **104:**165-172.

38. Kanzaki S, Beyer L, Karolyi IJ, Dolan DF, Fang Q, Probst FJ, Camper SA, Raphael Y: **Transgene correction maintains normal cochlear structure and function in 6-month-old Myo15a mutant mice.** *Hear Res* 2006, **214:**37-44.

39. Walsh T, Walsh V, Vreugde S, Hertzano R, Shahin H, Haika S, Lee MK, Kanaan M, King MC, Avraham KB: **From flies' eyes to our ears: mutations in a human class III myosin cause progressive non-syndromic hearing loss DFNB30.** *Proc Natl Acad Sci USA* 2002, **99:**7518-7523.

40. Albert S, Blons H, Jonard L, Feldmann D, Chauvin P, Loundon N, Sergent-Allaoui A, Houang M, Joannard A, Schmerber S, Delobel B, Leman J, Journel H, Catros H, Dollfus H, Eliot MM, David A, Calais C, Drouin-Garraud V, Obstoy MF, Tran Ba Huy P, Lacombe D, Duriez F, Francannet C, Bitoun P, Petit C, Garabedian EN, Couderc R, Marlin S, Denoyelle F: **SLC26A4 gene is frequently involved in nonsyndromic hearing impairment with enlarged vestibular aqueduct in Caucasian populations.** *Eur J Hum Genet* 2006, **14:**773-779.

41. Delmaghani S, del Castillo FJ, Michel V, Leibovici M, Aghaie A, Ron U, Van Laer L, Ben-Tal N, Van Camp G, Weil D, Langa F, Lathrop M, Avan P, Petit C: **Mutations in the gene encoding pejvakin, a newly identified protein of the afferent auditory pathway, cause DFNB59 auditory neuropathy.** *Nat Genet* 2006, **38:**770-778.

42. Cho KI, Lee JW, Kim KS, Lee EJ, Suh JG, Lee HJ, Kim HT, Hong SH, Chung WH, Chang KT, Hyun BH, Oh YS, Ryoo ZY: **Fine mapping of the circling (cir) gene on the distal portion of mouse chromosome 9.** *Comp Med* 2003, **53:**642-648.

43. Shabbir MI, Ahmed ZM, Khan SY, Riazuddin S, Waryah AM, Khan SN, Camps RD, Ghosh M, Kabra M, Belyantseva IA, Friedman TB, Riazuddin S: **Mutations of human TMHS cause recessively inherited non-syndromic hearing loss.** *J Med Genet* 2006, **43:**634-640.

44. Guipponi M, Vuagniaux G, Wattenhofer M, Shibuya K, Vazquez M, Dougherty L, Scamuffa N, Guida E, Okui M, Rossier C, Hancock M, Buchet K, Reymond A, Hummler E, Marzella PL, Kudoh J, Shimizu N, Scott HS, Antonarakis SE, Rossier BC: **The transmembrane serine protease (TMPRSS3) mutated in deafness DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro.** *Hum Mol Genet* 2002, **11:**2829-2836.

45. Rodriguez-Ballesteros M, del Castillo FJ, Martin Y, Moreno-Pelayo MA, Morera C, Prieto F, Marco J, Morant A, Gallo-Teran J, Morales-Angulo C, Navas C, Trinidad G, Tapia MC, Moreno F, Del Castillo I: **Auditory neuropathy in patients carrying mutations in the otoferlin gene (OTOF).** *Hum Mutat* 2003, **22:**451-456.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6947/7/32/prepub