Nucleic Acids Research, 1998, Vol. 26, No. 1 259–264

The Human PAX6 Mutation Database

Alastair Brown*, Mark McKie, Veronica van Heyningen and Jane Prosser

Medical Research Council Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received August 28, 1997; Revised and Accepted October 28, 1997

ABSTRACT

The Human PAX6 Mutation Database contains details of 94 mutations of the PAX6 gene. A Microsoft Access program is used by the Curator to store, update and search the database entries. Mutations can be entered directly by the Curator, or imported from submissions made via the World Wide Web. The PAX6 Mutation Database web page at URL http://www.hgu.mrc.ac.uk/Softdata/PAX6/ provides information about PAX6, as well as a fill-in form through which new mutations can be submitted to the Curator. A search facility allows remote users to query the database. A plain text format file of the data can be downloaded via the World Wide Web. The Curation program contains prior knowledge of the genetic code and of the PAX6 gene including cDNA sequence, location of intron/exon boundaries, and protein domains, so that the minimum of information need be provided by the submitter or Curator.

INTRODUCTION

The PAX6 Mutation Database was created to satisfy the need for a single source of information about human PAX6 gene mutations which are associated with developmental eye anomalies. It contains data produced in the MRC Human Genetic Unit, data gathered from the literature, and data submitted by researchers via the World Wide Web (WWW). PAX6 mutations have been found in a variety of dominantly inherited congenital eye defects including aniridia (absence of the iris), Peters anomaly, cataract, keratitis and isolated foveal hypoplasia. The overwhelming majority of PAX6 mutations have been found in aniridia patients, and these almost always lead to premature termination of translation of the PAX6 protein. Only four missense mutations have been reported to date; two associated with aniridia, one with Peters anomaly and one with foveal hypoplasia. A genotype–phenotype correlation is beginning to emerge: missense PAX6 mutations are very rare in typical aniridia cases and may tend to cause variant phenotypes. A review of PAX6 mutations based on the data contained in this database is in press (1).

The PAX6 Mutation Database system comprises two distinct parts. (i) The PAX6 Curation Program which is a Microsoft Access database providing facilities for the Curator to add/amend/search entries, to import new mutations submitted via the WWW, and to export the data for publication on the WWW. This program is only used by the Curator. (ii) A WWW site which contains HTML forms allowing remote users to submit new mutations and search existing data. Additional information such as gene maps and links to other PAX6 web sites is also included on the WWW page.

METHODS

PAX6 Curation program

The PAX6 Curation program was developed with Microsoft Access 2.0 (Microsoft Corporation) on a PC running Windows 3.1 and latterly Windows 95. A 32 bit Windows version for Windows 95/NT is being developed. Microsoft Access was chosen because the program is generally used only by the Curator and can easily be kept on a laptop computer.

World Wide Web Interface

HTML 3.2 compliant web pages and forms were written using Wordpad under Windows 95 or vi under UNIX. The HTML forms use JavaScript 1.0, but operate with non-JavaScript web browsers or have an alternative non-JavaScript version. All cgi scripts were written in C and compiled using the GNU C compiler (gcc, The Free Software Foundation). The CERN WWW server (httpd V3.0, available from http://www.w3.org/pub/WWW/Daemon/) was run on a Sun UltraSPARC 1 running Solaris 2.5.1 (Sun Microsystems Ltd).

WHAT IS IN THE DATABASE?

At the time of submission the database contained 94 mutations. Four of these describe neutral polymorphisms in the PAX6 gene which have been identified in normal individuals; the remainder describe mutations which have been found in individuals with aniridia and related eye disorders. Some of these mutations have been identified independently by different laboratories. ‘Compound’ mutations (e.g., deletion and insertion at a site) are treated as two or more mutations which are linked within the database; there are three examples of this to date. There are 35 database fields which are designed to provide as much information as possible about each sequence variant. Table 1 lists the database fields, the possible values and/or meaning of each, and whether the field is entered by the Curator/submitter or is one of 15 fields calculated by the Curation program using known information already stored in the database such as cDNA sequence, genetic code, nucleotide numbering of introns/exons (2) and location of domains. A few fields are optional, permitting additional information to be included. This particular database is specific to PAX6 but can be tailored to meet the requirements of other genes with differing structures.

*To whom correspondence should be addressed. Tel: +44 131 332 2471; Fax: +44 131 343 2620; Email: alastair.brown@hgu.mrc.ac.uk

Table 1. List of fields in the Human PAX6 Mutation Database with their meaning and possible values

DATABASE FIELD	POSSIBLE VALUES and/or MEANING	ENTERED BY SUBMITTER/CURATOR, or CALCULATED BY CURATION PROGRAM
PIN	>= 1 This is a unique numerical identifier in the database.	Calculated
COMPOUND MUTATION WITH PIN NO.	For compound mutations (e.g. deletion and insertion at a site) the PIN of the next "part" of the compound mutation. Each "part" is entered as separate item.	Calculated
LOCATION	Exon; Exon 5a; Intron; 5' Untranslated; 3' Untranslated;	Submitter/Curator
NUCLEOTIDE NUMBER	363 - 1631 for Exon; 1 - 42 for Exon 5a; any number for Intron; <= 362 for 5' Untranslated; >= 1 for 3' Untranslated	Submitter/Curator
CODON NUMBER	1 - 423	Calculated
EXON NUMBER	1 - 13	Calculated
DOMAIN	5'UT; PD; LNK; HD; PST; 3'UT;	Calculated
INTRON NUMBER	1 - 12	Submitter/Curator
MUTATION TYPE	Substitution; Insertion; Insertion with Partial Duplication; Insertion with Duplication; Deletion;	Submitter/Curator
TRANSITION OR TRANSVERSION	Transition; Transversion;	Calculated
ORIGINAL NUCLEOTIDE	C; G; T; A;	Submitter/Curator
MUTATED NUCLEOTIDE	C; G; T; A;	Submitter/Curator
ORIGINAL CODON	Any triplet (with left or right nucleotide shown in lower case if mutation at left or right end of codon).	Calculated
MUTATED CODON	Any triplet.	Calculated
ORIGINAL AMINO ACID	Any amino acid abbreviation.	Calculated
MUTATED AMINO ACID	Any amino acid abbreviation.	Calculated
INS/DEL SIZE	>= 1	Calculated
INS/DEL SEQUENCE	Sequence of insertion or deletion.	Submitter/Curator
SEQUENCE CONTEXT	Sequence around mutation site.	Calculated if mutation in codon
MUTATION SUMMARY IDENTIFIER	Identifier as per Ad Hoc Committee (1996) Human Mutation 8 197-202.	Submitter/Curator
PREDICTED RNA OUTCOME	Missense Sub.; Nonsense Sub.; Neutral Sub.; InFrameInsertion; FrameShiftInsertion; InFrameDeletion; FrameShiftDeletion; Probable Splice Error; Possible Splice Error; Unknown; or Unchanged;	Calculated for some mutation types
INFORMATION ON RNA OUTCOME	Any information supplied by the Submitter or Curator.	Optional
PREDICTED PROTEIN OUTCOME	Premature Termination; Out of Frame Extension; Internal Deletion; Amino Acid Substitution; Internal Insertion; Unknown; or Unchanged	Added by Curator (automatic determination may be used in future)
INFORMATION ON PROTEIN OUTCOME	Any information supplied by the Submitter or Curator.	Optional
PHENOTYPE	Aniridia; Congenital Cataracts; Isolated Foveal Hypoplasia; Autosomal Dominant Keratitis; Peters' Anomaly; Other; Normal (Polymorphisms); Unknown;	Submitter/Curator
OTHER RELATED PHENOTYPES	Any phenotype related to main phenotype.	Optional
OTHER UNRELATED PHENOTYPES	Any phenotype not related to the main phenotype.	Optional
INHERITANCE	Familial; Sporadic; Unknown; n/a	Submitter/Curator
MUTATION ID	User defined mutation identifier.	Optional
ORIGIN	Cell line; Tissue; Unknown; n/a	Submitter/Curator
SEX	Male; Female; Unknown; n/a	Submitter/Curator
REFERENCES	Citation; Unpublished;	Submitter/Curator
OTHER RELEVANT COMMENTS	Any comments made by submitter.	Optional
NEXT DUPLICATE	PIN Number of next mutation in database which is a duplicate of this one.	Calculated
DATE/TIME STAMP	Date and time of submission.	Calculated

Some fields must be entered by the submitter or Curator whilst others are calculated by the Curation program. Note that some fields are only relevant to some types of mutation.

PAX6 CURATION PROGRAM: MS ACCESS DATABASE

The PAX6 Curation program (Fig. 1) allows the Curator to add new mutations either directly, or by importing data submitted via the WWW form (see below). Data submitted via the web can be reviewed and checked by the Curator before final addition to the master database. The Curation program also allows searching of the data through a custom designed query form (Fig. 2). Data can be exported from the Curation system in various formats including delimited text for use as a flat file database, and Microsoft Excel format which can be converted to HTML for incorporation into the Web pages, used directly for publication (1) to produce multi-page tabular output, or to allow users to perform their own analysis of the data.

A demonstration version of the Curation program with read-only database can be down-loaded from the PAX6 Mutation Database web site. The program requires a 486 PC with 8 MB ram, 10 MB

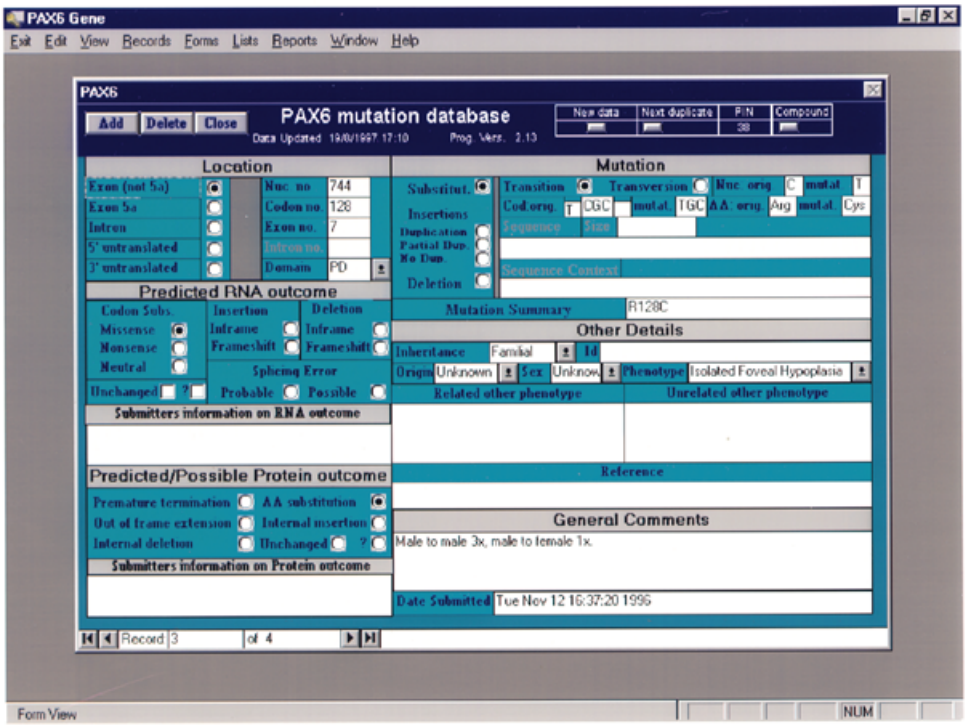

Figure 1. Screen from the PAX6 Curation program showing the record for a single mutation from the PAX6 Mutation Database. This is one of the records returned by the query shown in Figure 2.

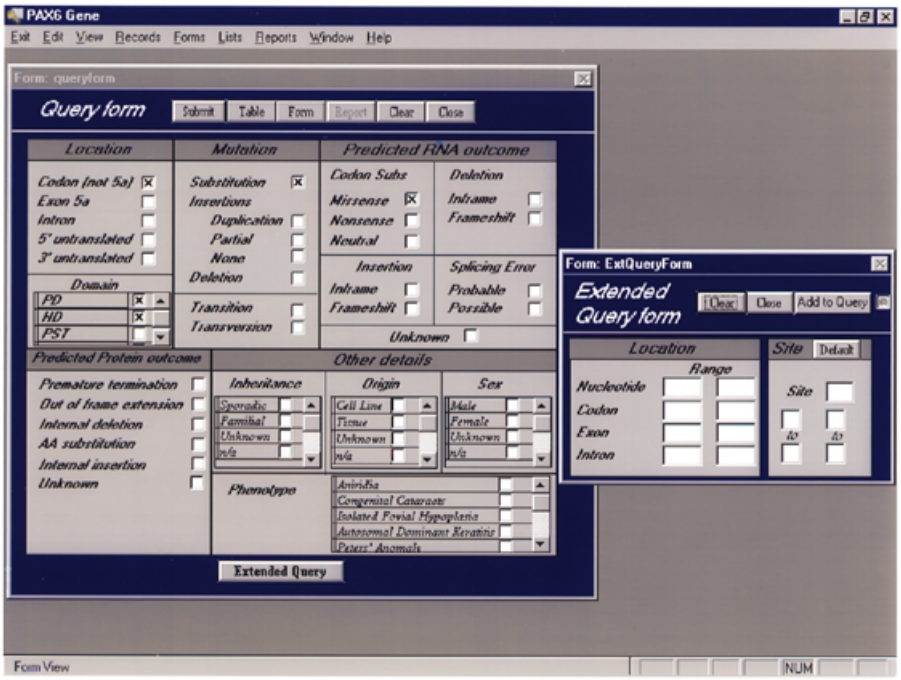

Figure 2. PAX6 query form from the PAX6 Curation program. The query is designed to return all missense mutations in the coding region covering the paired and homeo domains. One of the four records returned by this query is shown in Figure 1.

free disk space and Windows 3.x or above as a minimum configuration.

Importing data submitted via the WWW

A new PAX6 mutation submitted via the WWW is stored in a file in a secure network location for access by the PAX6 Curation program, and the PAX6 curator is automatically emailed with the name of the submitted file. Clicking the import button on the import/export screen of the Curation program causes the data to be read and converted into a Microsoft Access table of the same structure as the table storing the existing PAX6 mutation data. For each record in the import table, certain fields are calculated from existing information (see 'Adding a new record' below). The Curator can then choose to add each imported record to the master PAX6 mutation database and, if necessary, add further data or comments to each record.

Adding a new record

Adding a new mutation, either directly to the Curation Program or imported from a WWW submission, is made simpler by calculating the values of some fields. These are based on the location and type of mutation and the following pre-stored information: cDNA sequence, nucleotide numbering of exons and introns, location of domains, and the genetic code.

Data entry is broken up into discrete sections.

(i) Location. On indicating the region of the mutation, i.e., exon, intron, 5′ or 3′, some fields may be calculated automatically. If the region is exon, then entry of the nucleotide number at which the mutation occurs allows automatic calculation of the codon number, exon number, and domain where the mutation is sited.

(ii) Mutation. The type of mutation can be substitution, insertion or deletion. If a substitution, entering the nucleotide number allows calculation of the original nucleotide, codon and amino acid. When the substitution is entered, the program calculates the new codon, the new amino acid (if the mutation is not neutral), and whether the mutation is a transition or transversion. If an insertion, the Curator indicates whether it is an insertion with duplication, with partial duplication, or with no duplication. Entry of the insertion sequence causes automatic calculation of the size of the inserted sequence. If a deletion, then the deletion sequence is entered and the size calculated.

(iii) Predicted RNA outcome. From the type of mutation some predictions of RNA outcome can be made. If the mutation is exonic and a substitution, it can be determined if the resulting RNA will have a missense, nonsense or neutral outcome. If an insertion or deletion, it is possible to predict whether the mutation results in an in-frame or a frameshift outcome. The Curator can also indicate in this section if a splicing error is possible.

(iv) Predicted/possible protein outcome. Predictions of protein outcome are not made, but some limited predictions may be made in future versions. Currently the Curator predicts the outcome.

(v) Other details. This section is concerned with phenotype, sex, origin, inheritance, id and references. Details of possible values are contained in Table 1. There are also fields for submitter's experimental evidence for RNA and protein outcome, a comments box, and date of submission.

WORLD WIDE WEB INTERFACE

The PAX6 Mutation Database web pages (Fig. 3) contain general information about the PAX6 gene, but primarily provide (i) a means of submitting new mutations to the database, (ii) a means of searching the data, and (iii) links to other web sites containing information on the PAX6 gene.

Submitting new mutations

New mutations can be submitted to the Curator for addition to the database via a WWW form. The form comprises several sections which must be completed by the user. Items listed in Table 1 as being entered by the submitter must be completed in addition to name, organisation, address and Email address. Upon completion of the form, the user has the option of making a 'Test' submission which processes the data and performs error checks as normal, but displays the data on screen for the user rather than sending the data to the Curator. A variety of errors are checked, including missing data and inconsistent data such as invalid nucleotide number in the codon region or intron number outside the allowed range.

When the user is satisfied with the form data, they can be submitted to the Curator via the 'Submit' button. Upon submission, error checking is carried out (in case the user has not performed a 'Test' submission) and the data are stored in a file in a read-only archive directory of submissions. A copy of the file is made available to the Curator in another directory, and the Curator is automatically emailed with the file name of the new submission.

The Curator can then import the submission(s) into the PAX6 Curation Program where the data are checked before being added to the database proper (see above). If appropriate, the curator may contact the submitter to confirm or augment the details originally submitted, thus ensuring uniformity of the available information on each mutation.

Maximal information should be submitted, in particular any detail on phenotypic variation between family members and the presence of associated or unrelated anomalies. For example, if apparently unrelated phenotypic features recur at significant frequency, these may highlight previously unsuspected pleiotropic effects of gene function. This is the type of information which makes a mutation database an important tool in defining gene function in terms of phenotype.

Searching the system

The database can be queried by selecting options on a web form. The user can select the basic search query by checking the appropriate 'checkboxes' on the web form. Table 2 lists the groups of properties which can searched for in this way: items within groups are OR'd whilst properties between groups are AND'd. The search can be further narrowed by restricting the range of nucleotide number, codon number, exon number and intron number to be retrieved. Specific substitutions can be extracted, and sub-string searches can be included on text fields such as 'References' and 'Comments'. A choice of output is available: either full data for each matching record, concise data comprising selected fields, or mutation summary ID only (3).

Figure 3. The PAX6 Mutation Database web home page and contents list. The URL is http://www.hgu.mrc.ac.uk/Softdata/PAX6/

Table 2. Mutation properties which can be searched for through a 'point & click' World Wide Web interface

MUTATION PROPERTY GROUPS SEARCHABLE BY 'POINT & CLICK' INTERFACE
Location
Domain
Mutation type
Predicted RNA outcome
Predicted protein outcome
Inheritance
Origin
Sex
Phenotype

For possible values of each property see Table 1. The search can be narrowed further by searching for specific terms (see text).

Other data formats and demonstration Curation program

Should users wish to carry out a custom analysis of the data, the mutation database is exported in plain text (comma separated value) format. This file can be downloaded from the WWW page and imported into various packages such as Microsoft Excel. A demonstration version of the Curation program can also be downloaded.

Links to other web sites

Links to other PAX6 web sites include the PAX6 data at GDB, OMIM, and the GeneCard entry at the Weizmann Institute.

Links to other mutation web sites include the Human Gene Mutation Database, (Institute of Medical Genetics, Cardiff, UK), the Mutation Database Website (University of Melbourne, Australia) and Mutations at EBI (European Bioinformatics Institute, UK).

USING THE DATABASE

The PAX6 Mutation Database has been used extensively in the preparation of a review of PAX6 mutations (1). The database has proved useful in deriving statistics of the distribution of mutations across the gene (the paired domain has a noticeably higher level of mutation than other domains), and of the incidence with which specific mutations occur (most of the mutation in the homeodomain occurs in the hypermutable CpG dinucleotide in codon 240).

ELECTRONIC ADDRESSES

World Wide Web

The URL for the Human PAX6 Mutation Database is http://www. hgu.mrc.ac.uk/Softdata/PAX6/

Email

The Curator can be contacted by Email at pax6.curator@hgu. mrc.ac.uk

CITING THE HUMAN PAX6 MUTATION DATABASE

Users of the Human PAX6 Mutation Database are asked to cite this article and quote the above World Wide Web address.

ACKNOWLEDGEMENT

The authors would like to thank Isabel Hanson for her critical reading of the manuscript.

REFERENCES

1 Prosser, J. and van Heyningen,V. (1998) *Human Mut.*, In press.
2 Ton,C.C.T., Hirvonen,H., Miwa,H., Weil,M., Monaghan,P., Jordan,T., van Heyningen,V., Hastie,N.D., Meijers-Heijboer,H., Drechsler,M. *et al.* (1991) *Cell*, **67**, 1059–1074.
3 Ad Hoc Committee (1996) *Human Mut.*, **8**, 197–202.