

Bioinformatics tools for genetic services: a position paper on quality assessment of tools

Andrew Devereau, NGRL (Manchester)
Version 1.0
6 December 2005

Introduction

Project aims and background

This paper presents proposed approaches to the quality assessment of bioinformatics tools used by the genetics testing network in Europe. It forms part of work packages 2.2 and 2.3 within Unit 2 of the Eurogentest project, which seek to map the bioinformatics requirements of the genetic testing network and then to determine the quality of these tools and propose how they might be validated for use in diagnostics.

This paper has been developed as part of the initial survey of bioinformatics tools and requirements in order to inform the survey of genetic testing providers that will take place and to propose quality validation methods that can be used for the tools found. It has been developed from a literature survey.

In this paper I will summarise relevant data from the literature in which tools are categorised and quality assessment methods presented and propose how they might be applied to the categories of tools found. These proposals will be modified by data gained during a survey of genetic testers in Europe.

Scope

Eurogentest encompasses those disciplines involved in genetic testing, i.e. molecular genetics, cytogenetics and biochemical genetics. In the project we intend to survey the use of bioinformatics tools in representative laboratories throughout Europe. Which computer-based tools constitute bioinformatics applications will be one of the questions raised during the survey.

Development of bioinformatics tools

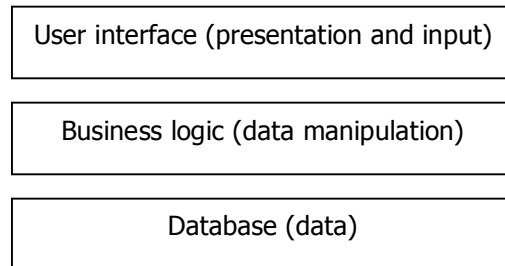
Reviews of the development of bioinformatics (Hagen 2000; Searls 2000; Ouzounis and Valencia 2003) suggest that techniques developed at the advent of computation biology in the 1970s still influence bioinformatics: analysis tools for sequence alignment, comparison of similar and homologous proteins and protein structure analysis were among the first tools to be developed; followed by development of databases for storage of sequences and results; and visualisation tools for presentation of data.

Searls (2000) suggests the following classification for bioinformatics tools:

- Databases and Data Resources
 - Database technology
 - Public databases

- Web resources
- Search and Analysis Tools
 - Similarity search and alignment
 - Pattern discovery and search
 - Gene finding
 - Gene expression
 - Genome annotation
 - Other tools
- Interfaces and visualisation tools.
 - GUIs
 - Scientific visualisation

I have adopted these categories as a useful starting point and suggest that quality measurement for each major category will be different. This categorisation is found in the other reviews (Hagen 2000; Ouzounis and Valencia 2003) and mirrors the layering found in typical software systems:



Categories of tools

Databases and data resources

Searls (2000) identifies databases as being at the heart of genomics, and Anon (2005) noted that it is the aim of databases to allow re-use of data between labs. In terms of technology most have moved from flat files to relational databases, and although there are some that have adopted object-oriented database technology this remains by far the most common platform used.

Searls (2000) describes the lifecycle of web resources. They typically start as an idea by researchers to fill a need and may then follow different paths: they may become disused (though the web site may persist); they may be continue to operate and be maintained by the original or like-minded people; they may attract funding for development and maintenance; and they may even become commercialised. These different states may themselves be an indication of quality in terms of how up-to-date or comprehensive the data are, and knowledge of who maintains and funds the database may also be important to potential users of the data: Searls (2000) notes that resources are offered virtually without any warranty other than that gained by inspection and knowledge of the developers' reputation.

There are many hundreds of databases available to the genetics community with a wide range of aims and scopes. A broad categorisation of such tools is: those presenting primary or raw data such as Ensembl¹, Genbank² and DDBJ³; those adding annotation to this data such as Flybase⁴

¹ www.ensembl.org

and SwissProt⁵; and those providing higher level structure and annotation such as Pfam⁶ (Birney, Clamp and Hubbard, 2002). Other types of large-scale data resource used in genetic testing include publication databases such as PubMed⁷ and disease and gene information resources such as OMIM⁸ and Orphanet⁹. Clearly there are overlaps between many of these and they have many different aims and scopes. Trying to categorise them to any detailed degree may be difficult and unproductive.

Many databases for genetic testing laboratories are aimed primarily at collecting and sharing mutation data. Within this more limited scope categorisation is more meaningful and is discussed by many authors. All describe core (or central) databases, which deal with mutations for many genes and diseases, and locus-specific databases (LSDBs), which deal with mutations for only one gene or disease, though Porter et al (2000) thought this categorisation an over-simplification. Patrinos and Brookes (2005) also describe National mutation databases, which hold mutation data for a particular population, and other resources such as SNP databases (dbSNP¹⁰, HapMap¹¹, PharmGKB¹²) and resources such as PhenomicsDB which holds multi-species phenotype-geneotype correlations. Some of the contrasting roles and features of core and LSDBs given in various papers can be summarised as follows:

Core (central) databases	LSDBs
More likely to be directly funded and therefore have full-time staff support for informatics and curation, and greater longevity.	Usually an additional project to the researchers' main interest so can become stagnant and may not have funding for expert input e.g. database/informatics support
There are those that validate all submissions, those that present all submitted data, and those that present only published data.	Will contain unpublished and published data – the aim is usually for completeness and high quality.
There are those that present each variation only once and those that present every instance.	Developers and curators are interested in the diseases so are interested in maintaining accurate up-to-date information, presenting a greater depth of knowledge and usually know submitters so are able to capture more data and be responsive to needs
May be infrequently updated, not interactive and have fewer or less flexible facilities such as search tools.	
No central database alone is sufficient for the needs of medical genetics: the approach is for	Typically aim to support a clinical or diagnostic service: the approach is for 'inch wide and

² www.ncbi.nlm.nih.gov

³ www.ddbj.nig.ac.jp

⁴ flybase.bio.indiana.edu

⁵ ca.expasy.org/sprot/

⁶ www.sanger.ac.uk/Software/Pfam

⁷ www.pubmed.gov

⁸ www.ncbi.nlm.nih.gov

⁹ www.orpha.net

¹⁰ www.ncbi.nlm.nih.gov

¹¹ www.hapmap.org

¹² www.pharmgkb.org

'mile wide and inch deep' coverage.	'mile deep' coverage.
UI will be consistent and there is the opportunity to carry out aggregated analyses across the data. There may not be the flexibility to represent data relevant to certain loci.	Content and quality extreme variability between genes.
Scale: HGMD (www.hgmd.cf.ac.uk) held 47889 mutations in 1885 genes on 02/09/2005	Typical LSDBs such as HbVar (globin.cse.psu.edu/globin/hbvar/), FXI (www.factorxi.com) and PAX6 (pax6.hgu.mrc.ac.uk) held 1234, 181 and 309 entries respectively on 11/11/2005.

References: Kalmar (2005), Claustres et al (2002), Scriver et al (1999) and (2000) Patrinos and Brookes (2005), Beroud et al (2005).

Claustres et al (2002) and Patrinos and Brookes (2005) noted that LSDBs and core databases have a similar function but are complementary because of their different depths, with each benefiting from the other. Cuticchia (2000) thought curation of central databases in detail is impossible, making it necessary for LSDBs to be established to provide this depth of information for a gene. Brown and McKie (2000) noted the same point: the advantages of a core database are a consistent UI and comprehensive dataset but the limitations are depth, and that they require considerable computer and human effort – this is overcome by LSDBs but only for specific genes. Scriver et al (1999) and (2000) produced a set of recommendations for the content, structure and deployment of mutation databases. To be classed as a knowledge base an LSDB must combine scientific and diagnostic data with information useful to clinicians or students, and information for patients and their families. They recommended that LSDBs use HGVS guidelines for content and nomenclature to reduced variability and noted that consortium operated sites have better potential viability and that it is desirable to maintain data security independently of curator's funding. Porter et al (2000) also noted that standardisation is a common issue and that LSDBs should use HUGO gene names¹³, HGVS nomenclature¹⁴ and RefSeq¹⁵ reference sequences.

Cotton and Horaitis (2000) discuss issues of databases and quality control, and treat the issue of data quality separately from that of the tools in question. Quality is equated to accuracy for the data and critical points in the laboratory and interpretation process are identified. Although such laboratory issues are probably outside the scope of this work it does highlight the need for such data to be collected and presented to the user to be able to judge the quality of data, and for validation of data where they have been interpreted e.g. aa residues. They recommend standardisation of the submission and review process for journals, using standardised data submission forms and tools for checking such as Mutation Checker from EBI. Thus the tools and facilities provided by a database or journal will affect the ability to implement quality control. They go on in Horaitis and Cotton (2004) to identify that data should be up-to-date and complete in mutation databases (data quality issues), with ready access and search facilities (tool quality issues). Some of the quality issues identified for mutation data were that nomenclature were non-uniform and central database curators were not expert in the gene. The LSDB generic system that they have developed (LSDB in-a-box) aims to be easy to install, platform independent, handles the core data identified as necessary for good quality and allows extension.

¹³ www.gene.ucl.ac.uk/nomenclature

¹⁴ www.hgvs.org

¹⁵ www.ncbi.nlm.nih.gov/RefSeq/

Birney, Clamp and Hubbard (2002) are also concerned with the quality of data: some of the issues that they raise include that data are most valuable when systematically organised and intergrated; that maintaining evidence trails is important so that derived data are linked to their evidence and changes are propagated appropriately; users need to know whether data are automatically derived, based on experimentation or curation and their accuracy, and there is an need to make measures of accuracy understandable in the context of their use.

Stenson et al (2003) describe the HGMD¹⁶ central database and this highlights some important quality issues regarding the scope and aim of databases and the policies adopted. For example, the types of mutation that this database holds are given, as is the coverage, the way that different mutations are represented, the policy of only recording each mutation once, the nomenclature standards used, the acceptance policy for data and details of the collaboration with their sponsors and how this affects the access to the data. Cuticchia (2000) also discusses the importance of making access policies and data sources clear along with the mission aims and funding. The importance of continued development and maintenance is stressed as is quality control of the development process and measures to guard against loss of expertise and data – more than one person has knowledge of the resource and it is documented and there are backup and archiving processes in place, and also that the project has long-term viability in terms of funding.

There are many descriptions of mutation databases in the literature: the journals Bioinformatics and Human Mutation have regular slots for databases. Many of these are for LSDBs or yearly updates from central databases. Recent examples include Beroud et al (2005), Brandon et al (2005), Hamosh et al (2005), Beysen et al (2005), Heinritz et al (2005), Leonard et al (2005), Tzoulaki et al (2005), Saunders et al (2005) and Scriver et al (2003). Some of the common factors that can be drawn from these papers which may be applicable to quality measurement include:

- Data source – published, submitted, collations
- Curation policy
- Provision of search tools
- Extent of linking
- Maintenance policy
- Scope of the data presented – what type of data is it
- How much data is there
- How fast is it growing
- When was it last updated
- How many users are there
- How does it overlap with other resources
- What are the limitations – e.g. OMIM does not have rearrangements
- What standards are used
- What is the funding basis and how long is this secured for
- What are the data security arrangements
- Who are the target users of the database
- What is it for
- Aim or mission
- Implementation – tools, methods etc.
- Nature of interface
- Nomenclature used
- Reference sequence used

¹⁶ www.hgmd.cf.ac.uk

- Data display methods – graphical, text
- Categories of mutations presented or how they are labelled.
- Architecture
- Qc procedure
- goals

Search and analysis tools

Searls (2000) discusses the range of tools that have been developed – standard tools include alignments, profiles, phylogenetic trees and gene finding, but there are many more. One problem with quality assessment is that there are many different types of tool each with a specific area of application. Even widely used tools such as BLAST have variants which are designed for specific applications, as well as new versions which aim to improve performance, and even though they may be based on the same algorithm or technique their quality may vary due to differences in programme design and the ability of the user to adjust and understand tool parameters. For these reasons it may be reasonable to suggest that one set of quality measures for such tools is a clear statement of what the tool is for, how it should be applied, what algorithm or approach it uses. Another set of measures can be based around performance measures of the tool – e.g. sensitivity and specificity, speed etc.

The European Bioinformatics Institute (EBI 2005) provides a directory listing a large range of tools and databases. The categories and instances of tools are:

[Similarity & Homology,](#)
[Blast2 - ASD,](#)
[Blast2 - EVEC,](#)
[Blast2 - NCBI,](#)
[Blast2 - Parasite,](#)
[Blast2 - WU,](#)
[Fasta,](#)
[Fasta - ASD,](#)
[Fasta - LGIC,](#)
[Fasta - Geno./Proteo.,](#)
[MPsrch,](#)
[Prot. Function. Analysis,](#)
[CluSTr,](#)
[GeneQuiz,](#)
[InterProScan,](#)
[Proteomic Services,](#)
[Dasty,](#)
[UniProt DAS,](#)
[Sequence Analysis,](#)
[Align,](#)
[ClustalW,](#)
[GeneWise,](#)
[PromoterWise,](#)
[Structural Analysis,](#)
[DALI,](#)
[DaliLite,](#)
[Maxsprout,](#)
[MSD Services,](#)
[MSDfold,](#)
[Tools Miscellaneous,](#)

[EMBL Computational Services](#),
[Expression Profiler](#),
[NEWT](#),
[QuickGO](#),
[Readseq](#),
[Web Services](#),

This shows how there are different versions of the most common tools like BLAST which are aimed at different situations and users, which in turn emphasises the need for users to be provided with information about the aim and use of each tool. EBI provides both a consistent user interface (UI) for the use of each tool and an extensive tool-specific help section.

NCBI-Blast2 Protein Database Query

BLAST stands for **B**asic **L**ocal **A**lignment **S**earch **T**ool. The emphasis of this tool is to find regions of sequence similarity, which will yield functional and evolutionary clues about the structure and function of your novel sequence. [WU-BLAST 2.0](#) and NCBI BLAST2 are distinctly different software packages, although they have a common lineage for some portions of their code, so the two packages do their work differently and obtain different results and offer different features. You can also check for vector contamination with [Blast2 EVEC](#).

YOUR EMAIL	SEARCH TITLE	RESULTS	DATABASE	PROGRAM
<input type="text"/>	Sequence	interactive ▾	Protein ▾ UniProt ▾	blastp ▾
ALIGN VIEWS	MATRIX	EXP.THR	FILTER	DROPOFF
pairwise ▾	blosum62 ▾	default ▾	false ▾	default ▾
OPENGAP	EXTENDGAP	GAPALIGN	SCORES	ALIGNMENTS
11 ▾	1 ▾	true ▾	default ▾	default ▾

Enter or Paste a PROTEIN ▾ Sequence in any format: Help

Upload a file: Browse... Run Blast Reset

EBI UI for NCBI Blast2

YOUR EMAIL	ALIGNMENT TITLE	RESULTS	ALIGNMENT	CPU MODE
<input type="text"/>	<input type="text" value="Sequence"/>	<input type="text" value="interactive"/> ▼	<input type="text" value="full"/> ▼	<input type="text" value="single"/> ▼
KTUP (WORD SIZE)	WINDOW LENGTH	SCORE TYPE	TOPDIAG	PAIRGAP
<input type="text" value="def"/> ▼	<input type="text" value="def"/> ▼	<input type="text" value="percent"/> ▼	<input type="text" value="def"/> ▼	<input type="text" value="def"/> ▼
MATRIX	GAP OPEN	END GAPS	GAP EXTENSION	GAP DISTANCES
<input type="text" value="def"/> ▼	<input type="text" value="def"/> ▼	<input type="text" value="def"/> ▼	<input type="text" value="def"/> ▼	<input type="text" value="def"/> ▼

OUTPUT		PHYLOGENETIC TREE		
OUTPUT FORMAT	OUTPUT ORDER	TREE TYPE	CORRECT DIST.	IGNORE GAPS
<input type="text" value="aln w/numbers"/> ▼	<input type="text" value="aligned"/> ▼	<input type="text" value="none"/> ▼	<input type="text" value="off"/> ▼	<input type="text" value="off"/> ▼

Enter or Paste a set of Sequences in any supported format:

Upload a file:

EBI UI for ClustalW

These UIs show several things: that the UI and the tool are separate things so that the quality of a tool can depend on which UI is chosen; that there will always be specific parameters for each tool type; that users will benefit from tailored help especially in choosing the correct parameters. The help provided by EBI in fact goes into the different BLAST tools and what each is for, allowing the user to choose the right one.

Our own recent experience of employing missense mutation prediction tools, although brief, supports this approach – we found difficulties firstly with understanding how to apply the tools, in terms of getting input data in the right format, understanding the options for using the tools and the parameters for their use. We then found it equally difficult to interpret the outputs in terms

of understanding whether they represented a correct application of the tool, and there was a lack of understanding of the reliance that could be placed on the results. This issue is addressed in the following paper.

Chavali et al (2005) address the quality of tools in terms of their performance. In their paper they discuss the relative performance of primer design tools, motivated by the apparent variability of T_m predictions by such tools and the ignorance of many users of the limitations of the software that they are using. Their approach was to develop a benchmark test that was applied to each tool with the results analysed statistically. These are presented with the aim of allowing users to employ tools with minimum deviation and to present the limitation and restrictions of tools to the users. In this latter respect the approach is like that of EBI (2005): inform the user to help them use the right tool in the best manner. But it is the provision of statistical information and benchmark testing that makes this paper unlike some others which present new tools or techniques. A brief inspection of some of the many new tools which are presented in journals each month shows that there are not always figures given for their sensitivity or specificity, or performance against benchmark tests. An example is Gu et al(2005) which describes a tool for graphical comparison of haplotype blocks: the details of the motivation for the work, the aims, the language, platform, algorithms, result reporting and UI are given and in these respects this is the same as the information that is required for databases. Ferrer-Costa et al (2005) do provide this information: for their tool for annotation of pathological mutations on proteins they include an assessment of accuracy, the training set used, the archetypes, what the output is and modes of operation. The training set and archetypes are important pieces of information which tell the users how the tool was developed and the gene that was the original target for development.

Interface and visualisation tools

These have developed in response to the large amount of data in genomes (Searls 2000). They include GUIs for data visualisation and interaction, which in some cases have developed into programming interfaces, toolkits and 'widgets'. They are separated from other tools in that they do not carry out analysis of data in themselves but present such data to users. Often however such interfaces will be incorporated into analysis tools and it may be difficult to separate them for quality assessment. Some of the quality issues will indeed be the same as those for other tools so it may be appropriate to deal with both analysis and visualisation tools in a similar way. One quality factor that may be especially relevant to this type of tool though is that of limitations – tools may have limitations to the extent of the data that they can display which should be made clear to the users to avoid the possibility that users assume data do not exist because they are not displayed.

Approaches to rating

In this section I will propose some different ways that the quality of databases and tools can be assessed or made assessable to the users.

Feature lists

Many of the papers presented above suggest that one way to make tools assessable to the users is to provide a clear statement of the features of a tool. This could include items such as who has developed the tool, how is it maintained, who funds it etc. Anon (2005) adds the project size and sustained funding in this category. Another set of measures could be based around the data, i.e. what is its provenance (or even is this data collected and given), how has its quality been assured, how up-to-date and complete is it etc. The approach of EBI is to make this type

of information available through the information that they make available about each tool or group of tools. In terms of databases the information could also include the depth of data in a database, i.e. contents may include the mutation alone, population specific allele frequencies, phenotype data. Other quality factors could concern the service or features of a database: tools for data submission, a credit and citation system, review for data validation and a browser for visualisation. Coordination and standardisation are important factors. These features may therefore be based around the features of the tool – standards used, contents, tools and features for submission and visualisation etc.

An analogy for this approach is that of the patient information given with medicines. These typically list, in an easy to understand and standardised format, what the name of the drug is, who makes it, what it is made from, what it is for, what it is NOT for, who should and should not take it, how it should be used etc. This information could be given in a similar way for any tool of database: what data it presents or what it is for, who maintains and funds it, how to use it, what method, algorithm or dataset it is based on, how to interpret the results etc. One important question is whether it should be the task of the tool developer to provide the information or whether it should be gathered by a third party as EBI do, and if the latter approach is taken whether it is practical to develop guidelines on how to use the tool.

A further analogy is given by hotel rating systems: these are based not on how good any hotel is but on how many of the required services or features are provided by each hotel. So, provided a hotel has a certain number of en-suite rooms, restaurant, 24-hour concierge etc. it can expect to receive a certain number of stars. It may be thought attractive to introduce a star rating system for bioinformatics tools: alternatively it may be more meaningful to provide a clear categorisation of tools and a list of the features provided. It is important to avoid the popular misinterpretation of star ratings as a subjective indication of quality.

A more applicable parallel may be drawn with the myGRID project (www.mygrid.org.uk). This is a bioinformatics project which seeks to build services that allow data and tool integration, allowing workflows or pipelines of bioinformatics tools to be assembled. One of the requirements of any tool that is to be used in such a service is that its input requirements and parameters are able to be discovered, and its outputs interpreted. There may be important lesson to be learned from this project.

Guttmacher (2001) discussed the difficulty in assessing the quality and accuracy of information given in web-based databases and related the issues to those addressed by ethics standards developed for health information websites. It is equally possible to provide excellent as it is to provide biased or misleading web content: it is suggested that basic questions are asked which include:

- Disclosure – does site disclose mission and ownership/support?
- Confidentiality – does it disclose its policy and is it sufficient to safeguard users
- Timely updating – are update frequent and label when this was done
- Expertise – does it list its staff and consultants? Is authorship of information clear? Do these people have appropriate expertise?
- Ethics codes – does it subscribe to any recognised code of conduct such as HON, Hi-Ethics principles or eHealth code of ethics.

These could be the basis of a features list for data. The ethical schemes mentioned both employ a set of principles that those who subscribe must abide by. HON (www.hon.ch) have eight principles which are: authority (the providers are medically trained or it is clear that they are not); complementarity (the info is to support not replace physicians); confidentiality (personal data confidential); attribution (source of data is clear); justifiable (claims supported by clear evidence); transparency of authorship (information is clear and contacts are given); transparency of sponsorship (support is clearly identified); honesty in advertising and editorial policy (use of advertising will be made clear). An EU project (EC 2002) is based on these principles. Hi Ethics (www.hi-ethics.org) has: Privacy policies; enhanced protection for health information; safeguarding consumer privacy, disclosure of ownership and financial sponsorship; identifying advertising and content sponsorship by third parties; quality of health information content; authorship and accountability; disclosure of source and validation for self assessment tools; professionalism; qualifications; transparency of interactions, candour and trustworthiness, disclosure of limitations; mechanism for feedback. Although these guidelines are designed for health web sites many of the principles are applicable to data providers for genetic testing and could be adapted into a list of features.

The objective of a feature list approach is clear – it should aim to provide all the information required for a potential user to choose the right tool or database, to apply it correctly and to interpret and report the results fully and correctly.

Subjective assessment

Under this heading I mean any attempt to present a review of tools from the point of view of a user. This could include both the features and presentation of the tool and the quality of its results. An analogy for this approach is that of the Michelin guide – the star rating used by this guide is based to an extent on the assessment of quality made by inspectors, although there is also an aspect of the features provided. Similar approaches are taken by consumer journals offering reviews of products such as cars, cameras, computers etc., although again it is worth mentioning that these journals usually provide features tables to allow comparison of products features as well as a subjective assessment of the less tangible quality aspects.

This approach would require an independent review system to be developed and a credible organisation assembled to make and publish reviews.

Objective assessment

Under this heading I would place approaches which build upon the subjective review by developing standardised statistical or benchmarking tests which will measure such statistics as sensitivity and specificity, standard deviations or errors, applicable ranges and other measures applicable to the specific use of each tool type. It may be the case that such measurements are made for a representative set of data, e.g. certain genes or genomic areas, or that the tool is developed for a certain gene (the archetype). It is especially important that this information is presented alongside any assessment.

This approach is compatible with the other approaches suggested above and I think is essential if bioinformatics tools are to be adopted and reported by medical geneticists. It is likely to be applicable to analysis tools rather than databases, for which details of coverage and currentness may well serve the same purpose but would not be amenable to statistical analysis.

Medical guidelines approach

Clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” (Institute of Medicine, quoted in Agrawal and Shiffman, 2001). The systematic nature of these guidelines has allowed ratings instruments to be developed to counter concerns about the standards of scientific evidence used in their development. Agrawal and Shiffman (2001) state that a guideline’s quality should be measured by a “prospective evaluation of its effectiveness in achieving its intended health outcomes”, but that this type of evaluation is lacking for most guidelines. Instead, an evaluation of the methodology used to develop the guideline and of the content of the resulting document is used. The three key principles in developing high quality guidelines are identified as: they should be multidisciplinary; they should be based on a systematic review of the literature; they should explicitly link their recommendations to the supporting evidence. The IOM proposed an assessment instrument evaluating eight attributes: clinical applicability/scope, clinical flexibility, reliability/reproducibility and validity are concerned with the substance of the guideline; while clarity, multidisciplinary process, scheduled review and documentation are concerned with the development process. Although this instrument was found too complex to implement it has been the basis of other rating instruments.

Although there has been criticism of the role of guidelines in the practice of medicine the way that their quality is measured may be a good example of systematic quality measurement. It is not a perfect analogy as bioinformatics tools are not systematically developed to assist in clinical care, and may be relied upon to provide information instead of guiding service provision, so a requirement for literature review for example may not be applicable. However the approach of separating the substance or outcome of the tool or guideline from its development is in common with the separation of the statistical analysis of a tool’s performance from its features that I have discussed above, and some of the attributes for each – scope, reliability and validity of outcomes; and clarity, scheduled review and documentation of the development process – reflect the features and measurements that I have suggested.

Conclusion

In conclusion, I would suggest that there are two aspects to quality assessment of tools that should be developed. Firstly the user needs information about the development process of the tool: what it does, what it is for, who developed it, how it will be supported, the development process used etc. I have found a range of data of this sort in relevant papers but it is unstructured and tends not to categorise the tools with others. The effects on quality of this information are manifold: the correct choice of tool and choice of input parameters is vital to the correct use of the tool, and may only be understood once the purpose and basis of a tool are understood; tools must be used correctly in order to obtain correct results; without information about the provenance of a tool its reliability can only be taken at face value; and the level of service offered to others, such as a diagnostic testing service, can only be as good as the level of service offered by the tools and techniques it uses so this information must be readily available. EBI provide a good example of categorisation of tools, which allows for the choice of the best technique available for a particular task, and add their own user interfaces and help files to each tool to ensure that the operation and basis of each one is known. The second aspect of quality is the assessment of the performance of a tool, e.g. in terms of sensitivity and specificity, or completeness. This is vital for the correct interpretation of the results obtained which in turn is vital for the confidence of the user in using and reporting the results.

How can these approaches be applied? The development of quality assessment methods for medical guidelines shows that instruments – systematic checklists based on agreed principles for

the development of valid guidelines – can be used to assess the quality of development and I would suggest that a similar approach may be taken for the assessment of development process. This may also be thought of as a 'feature list' approach in which features which are agreed to be important for quality are identified for each tool. Further to this though is the need to develop standardised or benchmark tests for assessment of the quality of performance – this may require standard test sets or scenarios to be developed. In both cases it is difficult to see how these tasks can be achieved without an independent and credible group being responsible for their application.

Bibliography

- Agrawal, A. and Shiffman, R.N. (2001). Evaluation of guideline quality using GEM_Q. *Medinfo 10(Pt 2):1097-101*
- Anon (2005). Editorial: WayStation to HUGOBase. *Nature Genetics 37(8):783*.
- Beroud, C, Hamround, D., Collod-Beroud, G., Boileau, C., Soussi, T and Claustres M. (2005). UMD (Universal Mutation Database): 2005 update. *Human Mutation 26(3):184-191*.
- Beysen D et al (2005). The human FOXL2 mutation database. *Human mutation 24:189-193*.
- Birney, E. et al. (2004) Ensembl 2004. *Nucleic Acids Research 32: D468-D470*.
- Birney, E., Clamp, M. and Hubbard, T. (2002). Databases and tools for browsing genomes. *Annual Review of Genomics and Human Genetics 3:293-310*.
- Brandon M.C. et al (2005). MITOMAP: a human mitochondrial genome database – 2004 update. *Nucleic Acids Research 33:D611-D613*.
- Brown A.F. and McKie M.M. (2000). MuStaR and other software for locus-specific mutation databases. *Human Mutation 15:76-85*.
- Chavali S et al (2005). Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics 21(20): 3918-3925*.
- Claustres et al (2002). Time for a Unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Research 12(5):680-688*.
- Cotton, R.G.H., Horaitis, O. (2000). Quality control in the discovery, reporting and recording of genomic variation. *Human Mutation 15:16-21*.
- Cuticchia A.J. (2000). Future vision of the GDB Human Genome Database. *Human Mutation 15:62-67*.
- EBI (2005). EBI Services. www.ebi.ac.uk/services and linked pages.
- EC (2002) eEurope 2002: Quality Criteria for Health Related websites. Brussels, 29/22/2002 Com (2002) 667 final.
- Ferrer-Costa et al (2005). PMUT: a web based tool for the annotation of pathological mutations on proteins. *Bioinformatics 21(14): 3176-3178*.

- Gu S., Pakstis, A J, Kidd, K K. (2005). HAPLOT: a graphical comparison of haplotype blocks, tagSNP sets and SNP variation for multiple populations. *Bioinformatics* 21(20):3938-3939.
- Guttmacher, A.E. (2001). Human Genetics on the Web. *Annual Review of Genomics and Human Genetics* 2:213-233.
- Hagen, J.B. (2000). The origins of bioinformatics. *Nature Review Genetics* 1:231-236.
- Hamosh A et al (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* 33:D514-D517.
- Health on the Net Foundation. www.hon.ch.
- Heinritz H et al (2005). The human TBX5 gene mutation database. *Human mutation Database in brief #846 Online*.
- Hi Ethics. www.hi-ethics.org.
- Hoffman M, Arnoldi, C. and Chuang, I (2005). The clinical bioinformatics ontology: a curated semantic network utilizing refseq information. *Pacific Symposium on Biocomputing 2005*:139-150.
- Horaitis, O and Cotton, R.G.H. (2004). The challenge of documenting mutation across the genome: the Human Genome Variation Society approach. *Human Mutation* 23:447-452.
- Kalmar L. et al (2005). HAEdb: a novel interactive, locus-specific mutation database for the C1 inhibitor gene. *Human mutation* 25:1-5.
- Leonard H et al (2005). Genotype and early development in Rett syndrome: the value of international data. *Brain and development (in press)*.
- Oetting WS and Tabone T (2005). The 2004 Human Genome Variation Society Scientific Meeting. *Human mutation* 26(2):160-163
- Ouzounis, C.A. and Valencia, A. (2003). Early bioinformatics: the birth of a discipline – a personal view. *Bioinformatics* 19(17):2176-2190.
- Patrinos G.P. and Brookes, A.J. (2005). DNA, diseases and databases: disastrously deficient. *Trends in genetics* 21(6): 333-338.
- Porter, C.J., Talbot, C.C. Jr., Cuticchia, A.J. (2000). Central Mutation Databases – A review. *Human Mutation* 15:36-44.
- Saunders R E et al (2005). Factor XI deficiency database: an interactive web database of mutations, phenotypes and structural analysis tools. *Human mutation*: 26:192-198.
- Scriver CR et al (2003). PAHdb 2003: what a locus-specific knowledgebase can do. *Human mutation* 21:333-344.
- Scriver CR, Nowacki PM and Lehvaslaiho H and the working group (2000). Guidelines and recommendations for content, structure and deployment of mutation databases: II. Journey. *Human Mutation* 13:344-350.

Scriver CR, Nowacki PM and Lehtsaio H (1999). Guidelines and recommendations for content, structure and deployment of mutation databases. *Human Mutation* 13:344-350.

Searls D.B. (2000). Bioinformatics tools for whole genomes. *Annual Review of Genomics and Human Genetics* 1:251-279.

Splendore A et al (2005). *TCOF1* mutation database: novel mutation in the alternatively spliced exon 6A and update in mutation nomenclature. *Human mutation* 25:429-434.

Staats, B. et al (2005). Genewindow: an interactive tools for the visulisation of genomic variation. *Nature Genetics* 37(2):109-110.

Stein, L.D. (2003). Integrating biological databases. *Nature Reviews Genetics* 4:337-345.

Stenson P.D. et al (2003). Human Gene Mutation Database (HGMD): 2003 Update. *Human Mutation* 21:577-581.

Syvänen, A-C, Taylor, G.R.T. (2004). Approaches for analysing human mutations and nucleotide sequence variation: a report from the seventh international mutation detections meeting, 2003. *Human Mutation* 23:401-405.

Tzoulaki I, White IMS and Hanson IM (2005). *PAX6* mutations: genotype-phenotype correlations. *BMC Genetics* 6:27.