Visualising A Skeletal Dysplasia Knowledgebase

Ingrid B Jakobsen1, 2, Theodor G Wyeld3, David P Hansen2, Andreas Zankl4

1 Queensland Facility for Advanced Bioinformatics, Australia
2 E-Health Research Centre, Queensland, Australia
3 Media, H&SS, The University of Adelaide, Australia
4 Royal Children’s Hospital, Brisbane, Australia
{Ingrid.Jakobsen@csiro.au, theodor.wyeld@adelaide.edu.au, David.Hansen@csiro.au, Andreas.Zankl@gmail.com }

Abstract

Skeletal dysplasias affect around 4 million people worldwide. Various nosologies exist for describing and classifying their clinical, radiological, and genetic features. A methodological framework is needed to establish a single consolidated nosology for skeletal dysplasias based on the existing multiple classification systems. Once established, the terms used can be linked to an existing database of exemplar x-ray images and explored interactively. As a proof-of-concept, a pilot study was conducted to investigate the potential in both establishing a consolidated nosology and visualisation of the results. This paper reports on a preliminary methodological framework for proceeding and initial visualisation results which go some way towards demonstrating the potential of such a system. The aim of this project is to assist clinicians to make more accurate diagnoses of skeletal dysplasias and to foster translational research. This is set against the background of an existing unwieldy classification system.

Keywords: 3D visual database, dysplasia, x-ray images,

1. Introduction

Worldwide there are estimated to be 4 million people suffering from skeletal dysplasias. This is a heterogeneous group of genetic disorders affecting skeletal development. It is a highly specialised area with less than 100 experts who specialise in skeletal dysplasias worldwide. Diagnoses are largely based on clinical examination of the patient and x-ray images of various skeletal regions. To-date over 370 conditions have been identified. Expensive gene testing can be used to confirm the initial clinical and radiological diagnosis. However, the molecular mechanisms are only understood for a minority of the diseases. Nosologies (the systematic classification of disorders) have been established for classifying skeletal dysplasias based on their clinical, radiological and genetic features. While clinical-radiographic classifications have dominated in the past, the increasing knowledge on the genetic basis of these conditions has led to a more prominent role for genetic classifications in recent years.

The various nosologies exist in printed form as lists, groupings, or tables of features. The inflexibility of the printed format and the lack of integration with clinical and radiographic data make these nosologies of little use in clinical practice. Various special interest groups have identified the need to consolidate the various nosologies into a single knowledgebase with a controlled vocabulary that could be used by all. This could then be linked to exemplar x-ray images. To-date, REAMS [1] is the only similar system available (it is a traditional database driven system with text query linked to a single nosology). Hence, there is a pressing need, firstly, to investigate ways of organising the various nosologies into a single knowledgebase, and secondly, to include exemplar x-ray images for comparison in this knowledgebase in an easily assimilable way.

In 1970, the International Working Group on the Classification of Constitutional Disorders of Bone devised the First International Nomenclature of Constitutional Bone Diseases [2]. This nomenclature has been revised on several occasions over the proceeding years. The increasing importance of molecular biology in the classification of skeletal dysplasias has led to a hybrid classification system incorporating clinical, radiological and molecular genetic criteria. In 2001 the Working Group identified the need to consolidate the various classification systems [3]. However, little has been achieved in the interim [4]. This paper outlines a method for not only organising the competing systems but uses a visualisation schema that assists clinicians in their everyday diagnostic practices.
the complexity using a visual database with multiple cases to compare.

This paper is organised into six sections: a background on skeletal dysplasias (SD), aims of a knowledgebase, existing information, defining the boundaries of the knowledgebase, components of a SD knowledgebase, and visualisation of a networked SD knowledgebase with associated x-ray imagery. It concludes with an overview of the need for a consolidated nosology and its visual linking to exemplar x-ray images, with recommendations for future research.

2. A Skeletal Dysplasia Knowledgebase

2.1. Skeletal Dysplasias

Skeletal dysplasias are congenital conditions of abnormal development of the skeleton. There is no exact definition of what constitutes a SD, as is the nature of biology. The disorders considered by SD experts are generally defined by having either primarily skeletal effects, or significant skeletal effects; mostly leading to shorter than normal stature; and having a genetic/heritable cause. In fact the disorders which are classified as SDs change over time; disorders previously considered distinct are merged, disorders given one name are split, disorders are no longer considered SDs, disorders are (newly) recognised and added, and other disorders are included.

Currently, about 300 distinct entities are recognised, many of them extremely rare. Determining the type and/or cause is important for treatment and counselling. Most diagnosis is based on examination of X-rays (radiology) with experts in the field pattern-matching a patient to their internal knowledge or reference sources.

More recently, the underlying genetic cause of many SDs have been determined, allowing diagnosis by genetic testing. This has also caused some re-examination of diagnoses and categories, as what were considered distinct disorders are found to be caused by the same gene. However, there are also examples of genes where different mutations cause distinctly different symptoms, either due to severity of effect, or to, for example, multifunctional proteins where some mutations may disrupt only some functions.

2.2. Aims of a Knowledge Base

2.2.1. Classification

Classification aids diagnosis, as one can quickly focus on the most likely candidate disorders for a given patient. Similarly, for a disorder where the gene is unknown, considering disorders with related phenotypes may provide target genes to test, either those already known, or related genes. Broader questions relate to understanding the genotype-phenotype relationship and the normal development of the skeleton. Geneticists are interested in how the genes interact, at what developmental stages, and how they function normally and when mutated.

2.2.2 Community Involvement

Currently, experts in diagnosing and classifying skeletal dysplasias from around the world meet every 4 years to discuss and agree on a classification, which is then published in a summary form (e.g. [3], [4]). There would be significant advantages if these experts could contribute to a knowledgebase: the information would be more up-to-date, far more detail could be readily available, and different classification approaches could be supported, rather than requiring a single classification such as the current compromise between clinical and genetic classification.

There is evidence from other areas that knowledgebases that are curated and updated by experts in the field are more useful and have longer functional lives – for example OMIM [5]. Nevertheless, many expert systems tend to “lock in” the data, distributing it on a CD-ROM for example. This is despite the fact that it is generally seen as advantageous if the data, and relationships among the data, are kept separate from the algorithms that are used to query the data.

2.3 Existing Information

2.3.1 OMIM

OMIM [5] is a valuable resource for genetic disorders due to its comprehensiveness, cross-references, and continual updating. Any skeletal dysplasia-specific knowledgebase would at the very least need to link to each corresponding OMIM entry to gain any acceptance in the user community. There are some significant disadvantages to simply importing the relevant subset of OMIM: the clinical and diagnostic information is somewhat limited and does not conform to a strict terminology. They are also not written by experts in the field. Additionally, OMIM entries tend to be updated by adding more text at the bottom of an entry, so entries become very long and it may take significant time to determine the current state of knowledge.

2.3.2 Genetic data

When classifying SDs based on gene function and pathways, it is desirable to use existing, generally-accepted databases and ontologies, particularly UniProt and GO (Gene Ontology). It is worth noting that these two are so well-accepted that most other databases, e.g. the Reactome, cross-reference using UniProt and GO accession numbers. Thus, a skeletal dysplasia knowledgebase can link to a wide variety of information provided the UniProt numbers for each gene is included.

There are many other databases such as GEO, the Gene Expression Omnibus, that could be linked and provide useful information for translational research on skeletal dysplasias, e.g. data mining for related disorders.
2.3.3 Clinical Terminologies

SNOMED CT is a clinical terminology published by the College of American Pathologists. SNOMED CT contains around 400,000 concepts, and over 1 million relationships between those concepts. The question becomes whether SNOMED CT is a suitable ontology for this project. Its advantage is that it is in use and is compatible with other projects. SNOMED CT’s shortcomings in the context of SDs are chiefly that it does not contain the truly specialised knowledge within this field, omitting many rarer and newer SDs, and also some of the specialised terminology for radiological features.

While the classification of SDs within SNOMED CT may not be useful for this project, the anatomical nomenclature and classification may prove more useful.

2.4 Defining the boundaries of the knowledgebase

In comparison to many other medical databases, one feature of the disorders which are classified as skeletal dysplasias is their relatively limited scope. This currently consists of about 370 SDs and the associated findings and genes. However, the boundary of the field of knowledge is hard to define. Firstly, there is no single definition of what constitutes a skeletal dysplasia. Many disorders have a mixture of skeletal and non-skeletal features. There are also several cases already known where the gene causing a skeletal dysplasia also causes other disorders. Similarly, anyone interested in the molecular mechanisms of bone formation will be interested in entire regulatory pathways of genes, even if some of those genes are also used in other systems and so lead to disorders not classified as skeletal dysplasias.

2.5 Components of a SD knowledgebase

A complete Skeletal Dysplasia knowledgebase would have a number of components. A list of the disorders, the clinical findings related to those disorders and the genes/mutations which cause them would form the basis of any such system. A controlled vocabulary and underlying ontology should be used to describe the clinical findings. A database of individual cases, with clinical findings associated with radiological images would provide exemplar information for the disorders.

Once this information is brought together, query forms for searching for diseases with a certain list of findings would aid clinicians in diagnosis. As shown below, the ability to visualise the images based on their relationships can greatly improve the experience of looking at the data. Other analysis of the data, using other classifiers such as Bayesian classifiers may also aid in diagnosis.

Finally, the community of SD experts should have the ability to annotate disorder and clinical information. This would enable the building of an active, currently accurate knowledgebase.

2.5.1 Database/knowledgebase structure

There are three primary categories of information required: individual clinical/radiological findings; diagnoses/syndromes; and genes. There are two primary classes of relationships among this information. Firstly, since most disorders are diagnosed based on a profile of findings, rather than a single unique diagnostic, the individual findings need to be able to form groups – features found together, or not. Thus, any given disorder would need to be linked to all the relevant findings. But, since some collections of findings are not recognised yet as disorders, or may represent a complex of disorders, findings also need to be able to be clustered independently of disorders (and ideally in the form of findings in each individual, or in a single radiograph). Secondly, the relationship implicit in the above – links between findings and disorders, findings and genes, and disorders and genes. The most important relationships in the database are therefore links between different types of information, or observed groupings of the same type of information (findings).

2.5.2 Ontologies

A controlled vocabulary is needed so that database search algorithms can find matches, and different terminology can be matched up to the same disorder or findings. The general approach in medical contexts is to allow lists of synonyms. It is rarer that the same term is used to mean different things in different contexts, and so needs its meaning restricted to one of those contexts.

The difference between a controlled vocabulary and an ontology is the existence of particular relationships between terms, in particular the “is_a” subsetting relationship, and to a lesser extent the “part_of” (possible) component relationship. These relationships are potentially valuable for at least some knowledgebase tasks.

An ontology, rather than merely a controlled vocabulary, of anatomical features and findings would allow better searching and reasoning – the ontology allows common knowledge about anatomy to be embedded in the database.

2.5.3 Text searching

Bayesian classifiers are the preferred tools for searching and pattern-matching. Significant advantages are that they can take strengths of links into account, and that they do not return a single yes-no answer, but instead a listing of possibility, ranked by score.

2.6 Visual Analysis of SD Imagery

Skeletal dysplasia diagnosis is an inherently visual process, clinically and radiographically – they both use images or sighting. Hence, using visual media to organise the classification system should help in the diagnosis process.
Current systems allow searching of lists of terms and features, such as REAMS [1] linked to a limited range of associated examples. The different specialist groups list terms according to what is important to them. This gives rise to ambiguities across the clinical, radiological, and molecular biological analyses. The consolidated list will be useful in ameliorating this ambiguity. Construction of the consolidated list would rely on contributions from the special interest groups to describe the various disorders thus creating a new over-arching nomenclature and more comprehensive collection of examples to work from. These can then be organised in a manner that makes it easier to compare similar features to those expressed across a group of x-rays. By providing a comprehensive graphical display of all possible disorders, clinical and radiographical features, and their genetic variations, the consulting physician can quickly focus on a narrow band of alternatives within a broad range of possible categories. Using a visualisation schema, they would not have to convert or translate their observations between their clinical images and exemplar images in the database. Such a system would not rely on any knowledge of specialist terminology to compare or extract the necessary information.

Instead of the existing hierarchical classification schemes, each skeletal dysplasia is annotated with a controlled vocabulary that describes the condition in the three dimensions: clinical features, radiographic features, and underlying molecular mechanism. In the prototype visualisation schema, scanned radiographs of classical examples of each condition are linked to the description. Investigation of visualisation principles was used to explore how these images can be displayed based on user-selected descriptors. Following a 3D visualisation schema, the images are visually grouped in disorders that share certain characteristics allowing the user to explore the radiographs across a group of related disorders. This provides a powerful diagnostic methodology that allows physicians unfamiliar with skeletal dysplasias to literally ‘zoom in’ on the relevant part of the nomenclature by specifying a few radiographic characteristics and then compare the radiographic examples with the radiographs of their patient to be diagnosed. Such a system would allow clustering of skeletal dysplasias by various parameters. Visually exploring such clusters may also help identify disorders that share a common pathogenic mechanism not yet identified under the existing system.

As a proof-of-concept, a small pilot application was implemented which links existing exemplar x-ray images to the nosology terms. It presents a navigable multi-dimensional data-space of images and terms. It leverages the power of the relational database and computer graphics with human pattern recognition [see 6]. Preliminary results indicate that greater efficiencies can be achieved in cross-checking common features between different SDs (see figure 1).

Conclusions

Skeletal dysplasias are not amenable to a single hierarchy of classification. Clinicians will most likely go through the process of categorizing and diagnosing skeletal dysplasias in different ways. This process will be heavily weighted by their experiences with patients they have previously treated. This can be especially problematic in the case of rare disorders.

To this end, we are developing a knowledgebase that allows different ways of searching and querying knowledge about skeletal dysplasias, to support individual practice. The use of a visualisation interface recognises the importance of visual analysis of radiological features present in x-ray images. It uses a comprehensive searchable database of exemplar x-ray images organised about their clinical, radiological, and genetic features. This addresses an important contemporary issue because there are so few clinical geneticists who specialise in skeletal dysplasias that some conditions may go undiagnosed. Indeed,
visualisation methods may make sense of the complexity, making it easier to identify existing conditions and identify new, undiscovered correspondences. Ideally, this will enable the experts to contribute new information into the knowledgebase.

The use of a knowledgebase which is maintained and annotated by an expert community will give a valuable resource both for the diagnosis and treatment of the disorders and for research into improving diagnosis and treatment, as well as understanding human skeletal development.

Acknowledgements

This project is jointly funded by the University of Queensland, the Royal Children’s Hospital Brisbane, the Queensland Facility for Advanced Bioinformatics, the E-Health Research Centre, and The University of Adelaide.

References


