IADIS International Journal on Computer Science and Information Systems
Vol. 11, No. 1, pp. 1-16
ISSN: 1646-3692

# AIDING GENETIC ANALYSTS: DESIGN OF A LITERATURE EVALUATION SYSTEM

Jorun Børsting. *Department of Informatics, University of Oslo, P.B. 1080, 0316 Oslo, Norway*.

Alma L. Culén. *Department of Informatics, University of Oslo, P.B. 1080, 0316 Oslo, Norway*.

Morten C. Eike. *Medical Genetics, Oslo University Hospital, P.B. 4950, 0424 Oslo, Norway.*

**ABSTRACT**

This paper is concerned with the design of a system that handles published research literature evaluation related to clinical DNA sequencing and analysis of genetic variants. The literature handling system is part of a larger system, the Norwegian clinical genetic Analysis Platform, currently under development at the Department of Medical Genetics at Oslo University Hospital. The genAP project has inquired into data handling requirements, procedures and supportive bioinformatics tools for analysis of genetic data. Finding and evaluating relevant literature that reports on clinical classifications of genetic variants is an important part of this process. In many cases, it is a requirement to compare local assessments with those published in high-quality external references, ensuring that the correct decision on the clinical nature of the variant is reached. The implications of the decisions made as part of this process are relevant for both patients and knowledge production and its transferability. We chose to use user-centered design as our research approach, in both qualitative (walk-troughs, interviews and talk-aloud evaluations) and quantitative (questionnaire) inquiries. User involvement in design and evaluation of the reference handling prototype was important for identifying diverse usability problems and design issues, which could then be improved in later iterations of the prototype. These issues included identifying the most relevant articles for a particular genetic variant and communicating uncertainty in individual assessments. Users have also contributed to defining more general guidelines for the re-design of later versions, e.g., a need for customization, as users often have different strategies for working with references. We assert that user involvement in the design and evaluation processes, such as described in this paper, leads to system design that is more in tune with users' needs, making the adoption and use of the system easier and improving the efficiency and quality of the analysis.

**KEYWORDS**

Usability, complex systems design, genetic sequencing, user-centered design, gene nomenclature.

# 1. INTRODUCTION

Genetic testing based on DNA sequencing is used in clinical practice for both diagnostic and prognostic purposes. Recent advances in the underlying technology, termed high-throughput sequencing (HTS), have resulted in vastly greater sequencing output for a fraction of the cost when compared to older techniques. This change has opened up for a massive increase in the clinical application of DNA sequencing, reaching more, and larger patient groups. HTS is therefore considered a crucial factor in making personalized medicine feasible. However, the enormous quantity of data generated by HTS and issues around knowledge extraction from that data are deeply connected with an increasingly important issue in bioinformatics, the handling of so-called "big data" (Schadt et al., 2010). At present, the analysis of DNA variants found in sequencing data involves a large and fragmented set of bioinformatics tools and informational resources, placing a high cognitive load on the individual analyst (Mardis, 2010). Although HTS technologies are effective in generating data, there is still a large developmental gap between sequencing output and final analysis results tailored to answer specific questions related to the genetic material (McKenna et al., 2010). Moreover, the lack of sophisticated and flexible applications that enable downstream analysts to access and manipulate massive sequencing data has been a hindrance to further development of tools and methods for sequencing (ibid.). Thus, it seems timely to look into ways of designing new systems based on research methods, tools and empirical findings from research fields such as human-computer interaction (HCI) and computer-supported collaborative work (CSCW). The approaches from these fields may help in design of systems that aid analysts in managing and interpreting data, focusing on their needs and their workflow, with an aim to reduce cognitive load and increase accuracy of the analysis (Eike et al., 2014).

One of the main problems with the usability of highly specialized systems, such as those used in DNA sequencing, is that highly qualified users are often not engaged in the design processes directly, resulting in systems that are not optimal for their use. The bioengineers, molecular biologists and physicians working with interpretation of results from DNA sequencing are users with high levels of expertise; they possess both tacit and complex domain knowledge, which are crucial for the analysis process. For the design of a clinical genetic analysis software to be successful, these users should be involved in the software development process, as has been argued by (Bolchini et al., 2009; Neri et al., 2012), among others. Accordingly, this paper focuses on user-centered design (Javahery et al., 2004), with user participation in both qualitative (walk-troughs, interviews, and talk-aloud evaluations) and quantitative (questionnaire) inquiries. The aim is to discover how genetic analysts work, what they do and how the new system could better support them in their work. A large number of possibilities for system improvement was identified and described in detail in (Børsting, 2014). A central tenet, crucial for the design of new systems for clinical genetic analysis, is to engage analysts in the design process and to include designers who, at least to some degree, understand the analysis processes.

In this paper, we focus on a small, but important, part of a new system interface developed as part of the Norwegian clinical genetic Analysis Platform (genAP) project, termed the *genAP interpreter*, at the Department of Medical Genetics at Oslo University Hospital. The genAP interpreter presents a structured, unified view of relevant information required to interpret genetic variants in a clinical setting, and guides the user through the interpretation process, as well as providing decision support (Eike et al., 2014). The part of the genAP

interpreter that is described here, is a reference evaluation module that enables analysts to handle relevant literature references to genetic variants. The main research questions addressed are: 1) How important are references for the analysis process? 2) How are references to variants handled today? 3) Are there some implications for the design of a new reference handling system that can ease the work or reduce the cognitive load for the analysts?

The results show that, while indeed very important for a decision process, references are handled in different ways by different analysts. Thus, rather than forcing the analysts to comply with the system, the new solution needed to provide some customization possibilities. Further, clear options for conveying uncertainty in assessments is necessary so that the next person looking at the same references may reach the same understandings. The identified issues, in conjunction with deeper understanding of existing practices around literature and its use in decision processes, provided guidelines for the design of a more successful handling of references in the re-designed system. Consequently, our third research question was answered in the positive. The present paper is an extended version of a conference paper (Børsting et al., 2015). The material added to this journal version is related to the evaluation of the prototype and the discussion of its future.

The paper is structured as follows: in the next section we provide more information on methods used during design and evaluation of the reference-handling system. In addition, we show why participants consider reference evaluation to be an important problem, and how it is performed today. Section 3 then addresses implications for the design that our data-gathering methods have yielded, showing the central issues. The prototype for a better reference handling procedures is then suggested and evaluated, also outlining its future development. Section 4 is dedicated to discussion of the results, followed by a short conclusion.

## 2. THE APPROACH AND THE DESIGN CONTEXT

In order to understand the context of the problem, we have studied the literature on the general workflow of genetic analysts, and usability problems that they experience with new sequencing interfaces. Several studies were found, such as that of Shyr et al., who state that *"a consensus opinion about a causal gene candidate may arise from a series of email exchanges, face-to-face meetings and sharing of references such as hyperlinks to scientific abstracts"* (Shyr et al., 2014, p. 134). The authors also point out that most software does not provide suitable functionalities for facilitating multiple users to collaborate on the same data, but that such software would be highly desirable and would accelerate the clinical analysis process. In our work, one of the first things we learned was that handling the literature references was one of the hardest things for analysts. The process had a collaborative, multi-user nature that was central, with a need for conveying assessments clearly, including any levels of uncertainty.

### 2.1 Method

In order to identify how users with high professional and domain knowledge actually work with literature related to genetic variants, we chose a user-centered design, with user participation. The methods chosen for the inquiry were both qualitative and quantitative, and are summarized in Table 1. The qualitative methods, such as observations and interviews, were used following the basic guidelines for user-centered design: 1) Focus on the user and

tasks from the start, 2) Involve users in the process of trying to find the solutions to the right problem, rather than solving a pre-defined problem in a better way, 3) When the right problem is identified, a solution, with users, may be iteratively improved (Gould and Lewis, 1985).

Table 1. Methods used to identify problems and guide design of the reference handling system

| Name | Focus | Method | Data gathered | Time | Participant |
|---|---|---|---|---|---|
| Identifying issues - User 1 | To identify the most challenging tasks and identify usability issues. | Talk aloud. Usability testing of prototype. Semi-structured interview. | Audio, video, pictures, notes. Document containing literature evaluation. | 1 hour 40min | 1 laboratory engineer |
| Identifying issues - User 2 | The same as for "Identifying issues user 1" In addition, to further explore and validate identified issues. | Talk aloud. Usability testing of prototype. Semi-structured interview. | Audio, video, notes and pictures. | 1 hour | 1 laboratory engineer |
| Observation 1 | To gather data about the reference evaluation functionalities. | Observation followed by a semi-structured interview. | Audio, notes and pictures. Variant classification documents. | 2 hours | 2 laboratory engineers |
| Observation 2 | The same as for the first observation. | Observation followed by a semi-structured interview. | Audio, notes and pictures. Variant classification documents. | 1 hour 12min | 1 laboratory engineer |
| Interview | The same as for the observations. | Semi-structured interview. | Audio recording and notes. | 1 hour 45min | 1 lab physician |
| Survey | The same as for the interview and observations. In addition, validate and further explore findings. | Survey sent out by e-mail to the future users of the system. | The Microsoft Word documents containing the survey answers. | | 11 participants |
| User evaluation | Perform a user evaluation of the prototype. | Prototype walkthrough. Semi-structured interview. | Audio, pictures and the prototype containing one literature evaluation. | 1 hour 30min | 1 laboratory engineer |

## 2.2 Case: Handling of Published Articles Referencing a Variant

Today, the process of analyzing gene variants and references is usually done consecutively by a minimum of three users. Typically, the procedure is as follows: a molecular biologist performs the initial analysis of observed gene variants, using different supporting software tools and external databases, as well as checking the existing literature for references to the observed variant. In this process, judgments are made based on general knowledge from molecular biology (such as the effect of a given variant on protein function) and genetics, but also based on literature references. The latter implies finding out whether conclusions about the clinical significance of a variant in question already exist, and are if the articles presenting the conclusions can be considered scientifically sound and trustworthy. The results of this work are then checked by another molecular biologist and, finally, by a lab physician. Usually only the first two users, but sometimes all three, comment on individual findings and articles and collaborate to determine the clinical classification of the gene variant, which describes the

clinical significance of the variant in standardized terms. In some cases, additional experts may be involved in the process. Although the analysts strive to reach conclusive decisions that describe the variant as either pathogenic (disease causing) or neutral, due to current knowledge limitations this is not always possible. In these cases, a classification category named Variants of Unknown clinical Significance (VUS) is used (Plon et al., 2008).

### 2.2.1 The Importance of Research Literature for the Analysis Process

Despite the existence of local and external databases with large collections of previously classified variants and associated references, variant classification practices can vary greatly between laboratories and over time, producing uncertainty whether presented findings are valid in a local context. Moreover, references found in external databases have often only been superficially evaluated, and includes many references of passing, and even lacking, relevance. Therefore, the literature still needs to be hand-curated. In addition, within genetics new research is published at a fast pace. It is therefore important to ensure that the latest research is taken into account and that the local database is properly updated (Dienstmann et al., 2014).

In our investigations, the first step was to identify main issues (Table 1) encountered in testing the new genAP interpreter prototype. A walk-through with two users was deployed, using the talk-aloud technique. This was followed by a semi-structured interview. The main finding from these user sessions was that the most difficult issue for analysts had to do with handling of literature references. Unpacking the meaning of 'difficult' is addressed next.

### 2.2.2 Understanding Literature Evaluation in Practice

Scientific research articles are reviewed and evaluated by analysts in order to determine if an observed gene variant is associated with the development of a hereditary disease. The analysts starts with a list of references to evaluate, which are usually obtained from various external sources such as the universal Human Gene Mutation Database (HGMD), gene-specific databases such as the Breast Cancer Information Core (BIC) and others that use the Leiden Open Variant Database (LOVD) system, as well as from manual Google and PubMed searches. When at least two independent articles are evaluated to be of high quality and with the same, high-confidence conclusion regarding the clinical significance of a gene variant, additional research references are often not further evaluated.

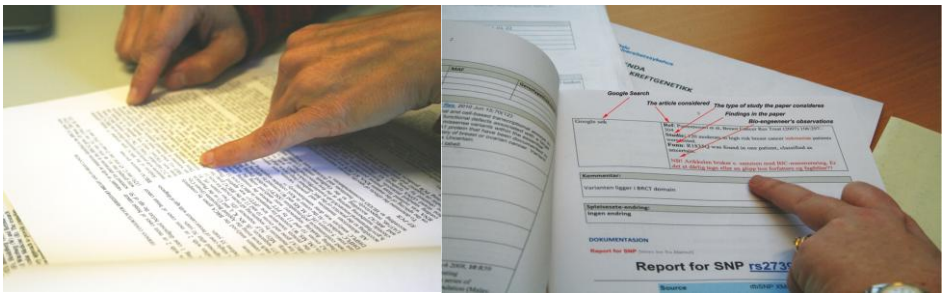


Figure 1. Observation of how the articles are handled showed that a Google search was performed, the selected paper printed, study type determined, results found, and then, in red ('NB! […]'), a note about uncertainty in findings (the trustworthiness of the paper) was written.

### 2.2.3 Data Gathering

Local workflow documents describing the handling of literature references were available, including departmental Standard Operating Procedures. Along with prior interviews with users, they served as the basis for the genAP interpreter prototype under development when this work started. As mentioned, our work started with the 'Identifying Issues' phase from Table 1, using observations. The purpose was to see if there is a difference between what users say (when interviewed about their work) and what they actually do (say-do problem, see (Kensing et al., 1998)), and to ensure that the correct set of problems were identified. Figure 1 shows how users start reading the paper and how they annotate it using the form that they have developed. Since the form is in Norwegian, our annotation in bold, with arrows, was added to the figure in order to explain different elements of the content. Findings from the interviews and further user studies in the form of a survey show that the analysts deploy different user strategies for handling challenges encountered during the variant classification process. Three main issues where users employed differing strategies were identified. These were not addressed in the local workflow documents or the first prototype, and were related to how the first article from the literature was chosen, how the individual analysts dealt with uncertainties regarding the trustworthiness of the article and, lastly, how they communicated their findings (to those evaluating the same variant later) in the comment field.

The first issue concerned cases where there is more than one relevant reference for a particular genetic variant. Since the analyst can stop evaluating new references when at least two articles meeting the requirements are found, we asserted that supporting user strategies that shorten the time spent on finding the right articles would increase the efficiency of the evaluation process. The answers from the survey show that users base their choice on different elements, some of which are shown in Figure 2.

| Question:<br>If you find more references for the current variant, how do you choose which one to evaluate first? | |
|---|---|
| Categories | Details |
| Publication. | Newest, tittle, author, journal, abstract. |
| Databases. | Google, HGMD, LOVD, BIC. |
| Variant related. | Relevance for variant and Variant mentioned in article. |
| Annotated articles | Descriptions from others. |
| Random choice | Any article as the start |
| Study type | Functional study, segregation analysis, family information. |

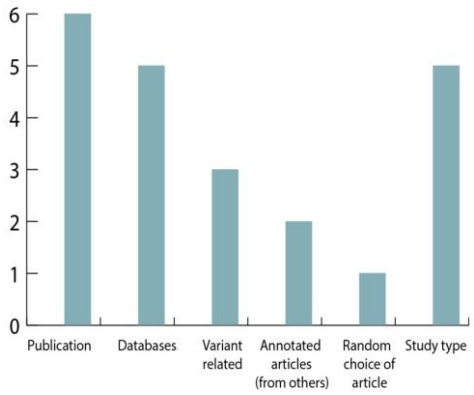


Figure 2. Results from the survey question regarding which articles that are evaluated first.

Direct observations of two users performing the article selection process showed that they searched the PDFs of each article to see how many times the variant in question was mentioned, before starting the actual evaluation. This, they stated, was a way to ensure that the variant was actually mentioned in the article explicitly, but also to get an initial 'intuitive' feeling about the article's relevance. In other words, this short search influenced whether the

article was considered a good candidate for being evaluated first. An observation was also that the name used for a particular variant is not always consistent in the literature. Thus, the analyst often had to perform multiple, manual searches, using different names for the variant in question. Based on these two observations, a suggestion for automatically providing the number of occurrences of a variant in an article was included in the survey. Answers to the question "Would this word-count be useful for you?" are displayed in Figure 3.

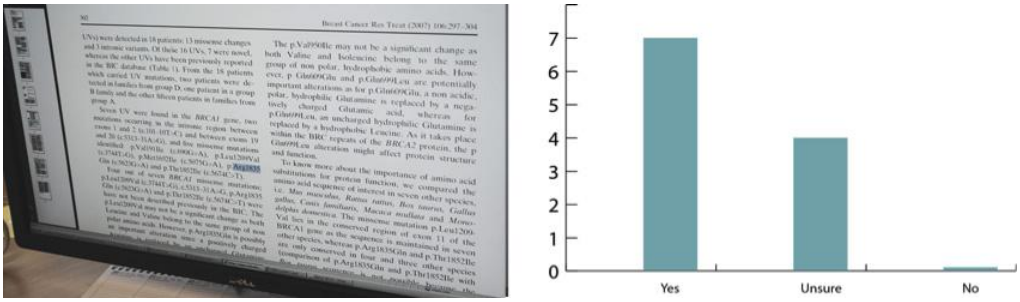


Figure 3. A genetic analyst searches for the variant in the PDF file of a potentially relevant article. The graph shows the answers to the survey question "Would this word-count be useful to you?"

Although this suggestion received a positive feedback, some users were unsure if the number of occurrences of the variant should be strongly correlated with usefulness of the article in the variant classification process.

The second major issue identified, and also the one where we observed the most variation in user strategies, was related to how users handled uncertainties in the assessment of a reference. One user discussed all such matters directly with a locally available colleague. In contrast, another user preferred to make her own assessments independently, and communicate via the comment field, the *level of uncertainty* in her judgment (if any). The next person doing the evaluation could then easily see this note and add a new assessment, or comment the previous one. The survey results also showed that several analysts were concerned about clarity of communication regarding the uncertainty in assessment processes. Some suggested the use of color-coding or highlighting the text containing uncertainty in the assessment.

The third major issue identified concerned how assessments are communicated to the next analyst via comment fields. As one analyst puts it: *"we copy and paste from the articles to convey to the next person that this was what we found. Then, the next person can find the places we copied from in the article and read it on their own* [note: assess and verify the content themselves]*"*. In the Survey the users where asked "What do you find important to include in the comment field?" The results related to this question are displayed in Figure 4.

Perhaps the most time-consuming part of evaluating articles relates to assessments of study quality. Some studies declare, for example, that they are functional studies, which is an important indication of a higher quality. However, authors' declaration is not enough. The analyst must check whether all procedures were done properly and assess if conclusions can be trusted. Also, the information pertaining to the specific variant under consideration is not always easy to extract from the article. For example, finding out what kind of study material (patient data, family history etc.) and method that was used on a particular variant is often difficult, as an article may include analyses of multiple variants using different methods.

Therefore, we conceded that the analysts needed a way to specify the study type and study material used for a specific variant. A comment field is suggested for this purpose.

| Categories | Users | Total |
|---|---|---|
| The article's conclusion regarding the variant | #1, #2, #3, #4, #5, #6, #10, #11 | 8 |
| Frequency data | #1, #3 | 2 |
| Information about patients | #1, #3, #4 | 3 |
| Splicing analyses | #1 | 1 |
| Description of study method | #1, #4 #9 | 3 |
| Functional study | #1, #3, #6, #9 | 4 |
| Family information | #1, #4 | 2 |
| Whether the article is well written | #1 | 1 |
| A summary of the article | #1 | 1 |
| Personal opinions about the article | #1, #3, #7, #10 | 4 |
| Date and name of prior evaluators | #2, #4 | 2 |
| Findings from articles referred to in the article | #5 | 1 |

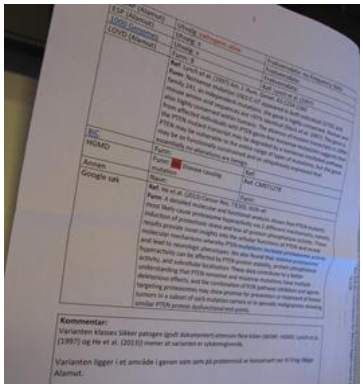

Figure 4. The table displays the results of the survey regarding the content of the comment field.

## 3. PROTOTYPING A NEW SOLUTION FOR REFERENCE HANDLING

Based on the analysis of all the data collected, a paper prototype was developed to further investigate our research questions. The prototype was a redesign of the genAP interpreter prototype mentioned above. Figures 5 and 6 show the new prototype. In Figure 5, a system provided list of articles to be evaluated is shown. An analyst needs to continue evaluating papers until at least two trustworthy articles are found for the same variant. Therefore, making good suggestions for sorting the papers and finding good articles faster reduces the overall time needed for classification of the observed variant.

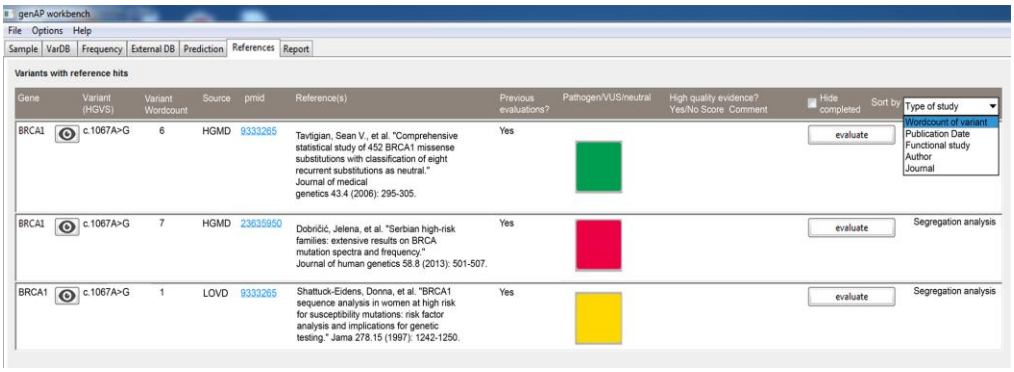


Figure 5. Prototype showing the list of references presented for analysts to evaluate. Two papers of high quality are considered as sufficient as input for making a decision.

The evaluated articles in the list are color-coded; red for pathogenic variants, green for neutral variants and yellow for VUS variants. These color choices were based on suggestions users made in the survey. The top navigation field is colored in a dark color with white text to increase visibility of the categories. The right-hand corner contains a drop-down menu for sorting the list by the type of the study, word-count, date of publication and other criteria that users mentioned as good selection practices.

Figure 6 shows the evaluation page that is displayed when an article is selected for evaluation (pressing the 'evaluate' button in Figure 5). On the top right-hand side, the variant word-count is displayed together with the particular variant name used in the article. The button labeled 'Find variant' can be used to find all occurrences of the variant in the text. To handle issues of uncertainty, the button 'mark with some level of uncertainty' is provided. When this button is used, the selected text in the comment field or in the article is highlighted (grey in this prototype). Parts of the user comments that are highlighted to showcase uncertain statements are removed when the variant classification is completed.

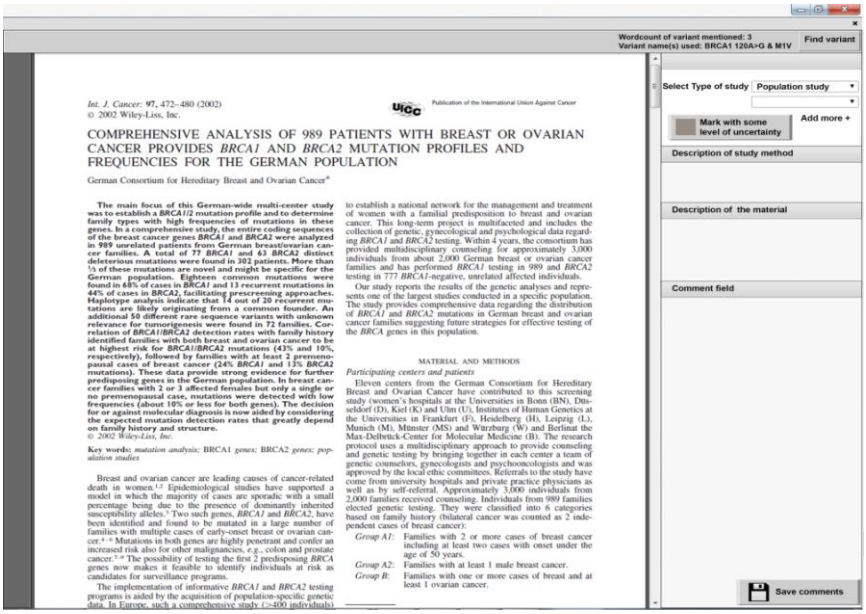


Figure 6. Evaluation page where the selected article is displayed, along with the comment fields used for its evaluation.

## 3.1 Improving the Prototype based on Users Feedback

Overall, what we learned through the design process, in particular observations and survey, was that it was of the outmost importance that the new system, ensures that assessments done by the first evaluator are clearly understood by the next person evaluating the same article. Secondly, the system needs to effectively support everyday work practices and provide higher efficiency. This is especially important for the genetic analysts we interviewed since, currently, the department is understaffed and the workload is steadily increasing. The user evaluation showed that the prototype was addressing both issues. Yet, it was opined that

further support and additional time saving functionality was still possible, and could be implemented in later versions of the system, preferably after a period of everyday use. This indicates that these users would like to be part of the changes made to the system, dynamically. Furthermore, we learned that tacit knowledge, based on which users develop a sort of 'intuition' aids them in article evaluation. For example, even if some article makes huge claims, the user 'intuitively knows' that such an article needs to be read more carefully, as these claims cannot be trusted a priori. When asked specifically how the system could support this important feeling of 'intuition', one user stated: *"*(When selecting the article) *the word-count and publication date are really useful* (to confirm the intuition). *Also, where the variant is mentioned in the article is important. If the variant is only mentioned in a table, then it´s not much we can use it for, other than to note that the variant is found before."* The evaluation, see Figure 7, also uncovered the five functionalities, presented in Table 2, that were perceived as particularly beneficial for the users and likely to be timesaving in their everyday work.

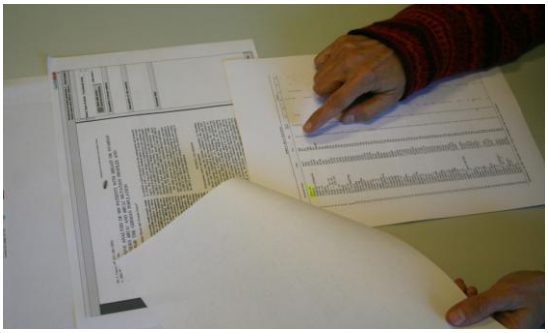

Figure 7. Walk-through session. The user was performing a set of pre-defined tasks.

Table 2. Most important findings from the evaluation of the prototype

| Important functionality | Explanation(s) |
|---|---|
| Displaying the particular name(s) of the gene variant used in the current article. | To eliminate the need for doing multiple searches using the different possible ways a variant may be named. |
| Presenting how many times the genetic variant is mentioned in the article. | To provide a quick initial impression on how much of the article the author has used to address the variant. This also eliminates the issue with articles that do not mention the gene variant. |
| A button that provides an automatic search of the variant name(s) used in the PDF. | To support the user strategy of traversing the PDF and to enhance efficiency, by providing automatic searches with the particular gene variant name used in the article. It is also important to include manual searches of the PDF, since there might be additional things users are looking for. |
| Providing the possibility for clicking on pasted text in the comment field and then to be guided directly to the exact location in the PDF, where the text was extracted from. | To support the communications between different users performing the article evaluation, by quickly displaying the article statements that are the basis for previous assessments. Since different users may assess the article statements differently, it is important that all evaluators read these statements in the article and form their own opinion. |
| The article's supplementary data files should be easily accessed within the program. | To eliminate the time used searching for this data online, which is both stated as time consuming and something that has to be done repetitively for the different variant classification cases, when various gene variants are addressed in the same article. |

Based on the findings from this walkthrough evaluation, the prototype was altered as shown in Figures 8 and 9. Changes included removing the PubMed-id ('pmid' column in Figure 5), since it was perceived as not relevant. Also, particularly good articles describing a specific gene more generally and, thus, often used, needed to be added to the article list for all variants within that gene. This was based on a suggestion from the user, who referred to such articles as a pool of 'special articles' that are added based on strict criteria specified by a super-user. In Figure 9, the comment field is also divided into two instead of three parts. The reason for this change is that the study method was already covered by the 'select type of study' dropdown menu, just above the comment field.

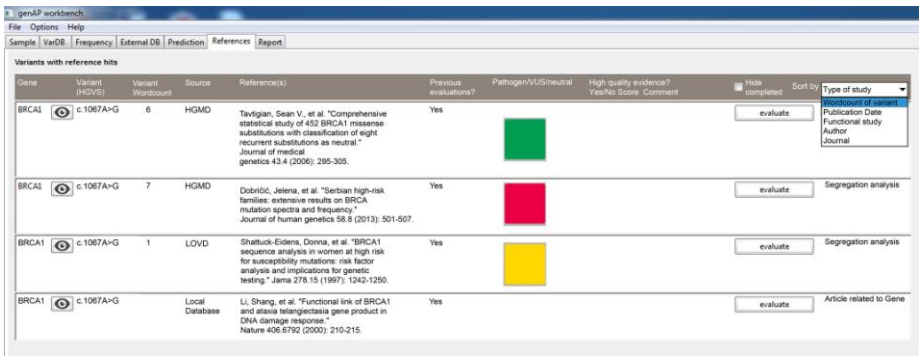


Figure 8. The changes in the prototype.

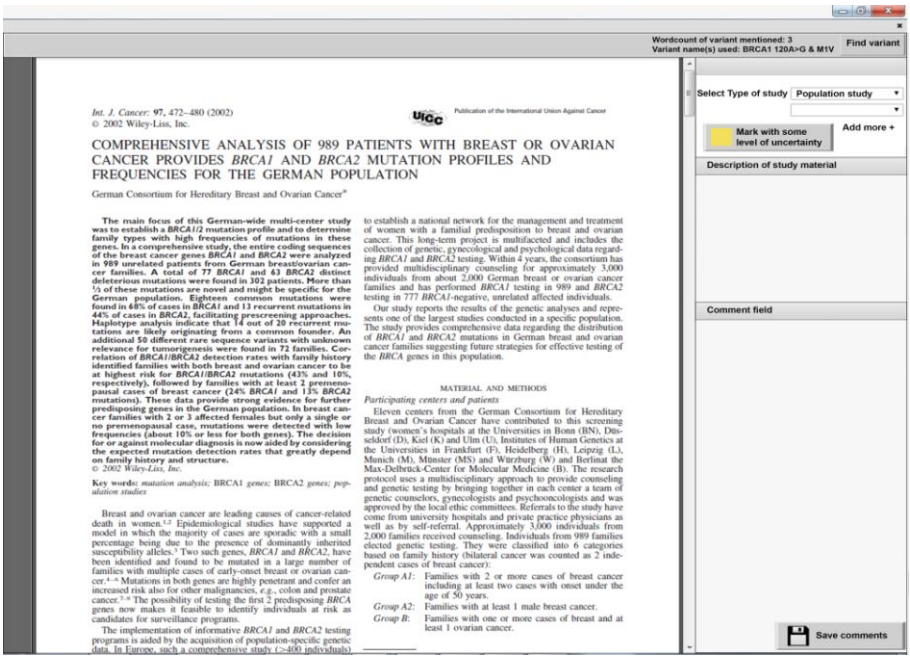


Figure 9. The article page changes: two fields for comments and yellow highlights.

Within genetics, sequence variant nomenclature is the scientific naming of genetic variants in relation to a particular context. In later years, universally agreed upon guidelines for this have been provided by the Human Genome Variation Society. However, these have changed over time, and there are still multiple ways of describing a variant within these guidelines. E.g., a variant may be named in relation to a chromosome, transcript or protein reference sequence. This inconsistency in the naming of genetic variants, pose particular problems for the users during the article evaluation addressed in this article. We found that by displaying the particular name(s) of the gene variant used in the article, the time used to finding the name(s) are eliminated. In addition, to ensure that all the correct names are counted, it is important that new versions of the sequence variant nomenclatures are added as they are put into use, so that the functionality stays up to date. One user expressed the perceived usefulness of the word-count functionality by stating *"If you see that the variant is mentioned many times, then you are pretty sure that this is a useful article to read."* Furthermore, the word-count would be used to select which article to evaluate first. When the user performed an article evaluation with the prototype, the following statement was made *"the number of times the variant is mentioned in the article and the publication date will be significant factors for me to choose which article I read first. In addition, to what type of study it is and which journal."* The user also considered the date of publication as important. Based on her experience, she knows that older articles are often a bit more vague when they make claims about their findings.

The last user evaluation uncovered two issues that were not currently covered by the prototype. The first was how users often found that articles listed for evaluation was not relevant, based on how they simply referred to another article. Often the later publication contains neither essential new findings nor more detailed descriptions of study method or material. In these cases, all the relevant and useful information was in the first publication. When this is found to be the case, the article evaluation is stopped and instead time is used searching for the original references. This could be avoided if the system had the ability to detect and communicate to the user that this is an article that only refers to older articles and has no valuable new findings.

The second issue was that if the gene variant is found many times in the article in combination with the word prediction, it is very likely that the article is not useful in the classification of the variant. Since if an article's conclusions are based solely on bioinformatic prediction tools, it is very likely not useful for classifying a variant, as these tools generally are not trusted. It could therefore be beneficial to broaden the functionality related to the search of article content to also include other keywords, in this case the word 'prediction', in association with the variant name. On the other hand, if the word 'mRNA' or other keywords that indicate the use of functional studies are found together with the variant name, then the likelihood is greater that the article contains information that is relevant for the variant classification.

As mentioned, the results in this study suggest that the design of our prototype provide timesaving functionalities and supports the communications of assessments between different users performing the article evaluation. Further support for such an understanding between the users and additional time saving changes should be addresses in later versions of the system if it is put into everyday use and practice. One example stated by the user is that how the comment is formulated will be established through use and that frequent phrases will be made. Functionality providing easy access to such phrases in the formulation of the comment could further increase efficiency. Even how the comment field is used will develop through time and new issues that should be addressed in later versions of the system could arise.

## 3.2 The Future of the Prototype

Since the first interviews, the development of the genAP interpreter has taken a new turn. In March 2015, the American College of Medical Genetics and Genomics (ACMG) issued new recommendations for how to interpret genetic variants in a clinical setting (Richards et al., 2015). These guidelines provide clearer criteria for how to interpret individual pieces of information, including those retrieved from literature references, and have rapidly been incorporated into the Standard Operating Procedures at the Department of Medical Genetics. This represents quite a large change in procedures, which also has implications for the design of the reference evaluation module. Based on these criteria, a new "rules engine" has been developed for the genAP interpreter, taking as input evidence that is categorized and weighted according to the ACMG guidelines, and providing a suggested clinical classification based on the sum of this evidence. The reference evaluation system has also very recently been redesigned to incorporate these changes, most importantly including a redesigned, buttoned evaluation form that outputs relevant ACMG-categorized information to the rules engine and displays them to the user. This means that the free form comments from the users are complemented by the structured output of the evaluation form. In the new version, "Type of study" and clinical classification (pathogenic/VUS/neutral) are also incorporated as user choices in the evaluation form.

However, the free form comments are still important, and are a central feature of the new design. Also, the functionality for generating alternative variant names have already been implemented, forming the basis for a word-count and search function as described in the prototype here. This function will therefore likely be implemented in one of the next versions of the reference evaluation module. Extending this to contextual searches using additional keywords, as suggested in the last user session and in (Børsting, 2014), as well as adding the suggested function for marking uncertain passages in the user comments and PDFs, are also currently under evaluation.

# 4. DISCUSSION

This study highlights the importance of providing system support for multiple user strategies when handling the literature findings related to classification of variants in genetic analysis. The strength of the work lies in drawing upon knowledge of genetic analysts and lab doctors through user involvement in the re-design process. From research described in the Table 1, it was evident that users valued that the system supported their workflow. This is in line with findings from (Shyr et al., 2014).

User involvement in the development of clinical decision support tools is also important, since the local work practices are often unique. Lindgren argues: *"the organization of clinical practice differs between clinics and countries. Local routines, work division, amount and characteristics of teamwork, etc., affect who may benefit from the support provided by a clinical decision support. Such factors need also to be taken into account when the user environment is assessed, and requirements for a CDSS* (Clinical Decision Support System) *are formulated"*, (Lindgren, 2011, p. 129). Our research also finds numerous characteristics and examples of local work practices and how the system can benefit from understanding and supporting those practices. Collaboration in the form of verbal discussions could also have

been a part of the system, but is not currently implemented. Many users did not see the benefit of supporting online discussions since they work in close physical proximity and have ample opportunities for face-to-face interactions. Others preferred not to have direct communication during evaluation of references, and adopted strategies such as the one we mentioned earlier, namely highlighting uncertainties in the text or comments in color. This example could be understood as an awareness-making mechanism and could be included into the new system as a support for collaboration. The users were generally positive towards such collaborative support. This is consistent with the finding in Shyr et al. that *"users expressed that an ideal system would allow users to attach notes, links to scholarly articles, as well as comments on individual genes or genetic variations, and that such information be available to multiple users in the same clinical setting. Software that empowers collaborative analysis would be well received"* (Shyr et al., 2014, p. 134).

Users handle different genetic variant classification cases by deploying diverse user strategies. These strategies need to be reflected in the design of the system in order to present the right information to the user at the time of decision-making and make their work less time-consuming. Our prototype incorporates strategies that were adaptable both to individual user preferences and work styles, and those brought on by demands from variant classification cases. As DNA sequencing technology and its uses is advancing and increasing the workload of genetic analysts, most likely the user strategies will change. One of the users addressed this by stating that the "*system has to be flexible*." This is again consistent with Shyr et al., warning that "*there are unique cases, which require unusual analysis approaches. Therefore while the software should be structured around specific standard analysis models, it needs to remain flexible*" (Shyr et al., 2014, p. 134).

Observing what users do, rather than just collecting data from interviews or surveys, was important. For instance, without observing users during the actual analyses, some findings would have been missed, as the users were not always able to articulate precisely what it is they actually do. What they said they did, and what they actually did were therefore in some cases different, representing a classic say/do problem (Simonsen and Kensing, 1998).

During the course of this study, we focused systematically on applying the user-centered design approach and its methods. These were an aid in maneuvering the complex research domain of genetic analysis, workflow and evaluation of literature references. The use of the approach helped discover the large amount of usability issues and shape them into a more flexible and user-friendly system. The identification of recurring design issues and themes were not done in order to make generalizations and force all users to work in the same way, but rather to explore how to support highly qualified individual users/bioengineers to work most effectively and based on their own tacit knowledge. We hope that the results we present demonstrate the benefits of taking user-centered approach also in the complex domain of bioinformatics.

## 5. CONCLUSION

The findings of this study indicate that user-centered design can be a good way of overcoming some usability challenges when working in complex domains. By including users, issues related to human-to-human interactions and collaborations also become visible. Thus, the chances of designing a system that provides wider and better support for analysts increases.

The application of user-centered methods revealed how users contributed with valuable input for the design of the future system. Such rich input could hardly be gathered in other ways, e.g., studying workflow charts. Understanding the ecology of the system and all the relations between technology and people needed to be considered and understood. Placing the analysts in the center, however, helped to adjust the focus on human productivity regarding the support for accuracy and speed of assessment. The case of reference management hopefully illustrates well these points.

## ACKNOWLEDGEMENTS

This research was part of the genAP project (Norwegian clinical genetic Analysis Platform), supported by The Research Council of Norway (grant no. 210622/O70). Our thanks are due to all lab engineers and physicians at the Department of Medical Genetics, Oslo University Hospital who participated in this study.

## REFERENCES

Bolchini, D., Finkelstein, A., Perrone, V., Nagl, S., 2009. Better bioinformatics through usability analysis. Bioinformatics 25, 406–412. doi:10.1093/bioinformatics/btn633

Børsting, J., 2014. Design of Genetic Classification Software: The Case of Representation of Research References.

Børsting, J., Culén, A.L., Eike, M.C., 2015. Design of a Reference Handling System for Clinical DNA Sequencing Analysis., in: Proceedings of the International Conference on E-Health 2015. Presented at the Proceedings of the International Conference on e-Health 2015, IADIS Press, pp. 79–87.

Dienstmann, R., Dong, F., Borger, D., Dias-Santagata, D., Ellisen, L.W., Le, L.P., Iafrate, A.J., 2014. Standardized decision support in next generation sequencing reports of somatic cancer variants. Mol. Oncol. 8, 859–873. doi:10.1016/j.molonc.2014.03.021

Eike, M.C., Skorve, E., Håndstad, T., Fontenelle, H., Børsting, J., Aanestad, M., Culén, A.L., Grünfeld, T., Undlien, D.E., 2014. GenAP workbench: aiding variant classification in clinical diagnostic settings, in: American Society of Human Genetics Annual Meeting. Presented at the American Society of Human Genetics, San Diego.

Gould, J.D., Lewis, C., 1985. Designing for Usability: Key Principles and What Designers Think. Commun ACM 28, 300–311. doi:10.1145/3166.3170

Javahery, H., Seffah, A., Radhakrishnan, T., 2004. Beyond Power: Making Bioinformatics Tools User-centered. Commun ACM 47, 58–63. doi:10.1145/1029496.1029527

Kensing, F., Simonsen, J., Bodker, K., 1998. MUST: A Method for Participatory Design. Human–Computer Interact. 13, 167–198. doi:10.1207/s15327051hci1302_3

Lindgren, H., 2011. Towards personalized decision support in the dementia domain based on clinical practice guidelines. User Model. User-Adapt. Interact. 21, 377–406. doi:10.1007/s11257-010-9090-4

Mardis, E.R., 2010. The $1,000 genome, the $100,000 analysis? Genome Med. 2, 84. doi:10.1186/gm205

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. doi:10.1101/gr.107524.110

Neri, P.M., Pollard, S.E., Volk, L.A., Newmark, L.P., Varugheese, M., Baxter, S., Aronson, S.J., Rehm, H.L., Bates, D.W., 2012. Usability of a novel clinician interface for genetic results. J. Biomed. Inform. 45, 950–957. doi:10.1016/j.jbi.2012.03.007

Plon, S.E., Eccles, D.M., Easton, D., Foulkes, W.D., Genuardi, M., Greenblatt, M.S., Hogervorst, F.B.L., Hoogerbrugge, N., Spurdle, A.B., Tavtigian, S.V., 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum. Mutat. 29, 1282–1291. doi:10.1002/humu.20880

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17(5), 405–425. doi:10.1038/gim.2015.30

Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., Nolan, G.P., 2010. Computational solutions to large-scale data management and analysis. Nat. Rev. Genet. 11, 647–657. doi:10.1038/nrg2857

Shyr, C., Kushniruk, A., Wasserman, W.W., 2014. Usability study of clinical exome analysis software: Top lessons learned and recommendations. J. Biomed. Inform. 51, 129–136. doi:10.1016/j.jbi.2014.05.004

Simonsen, J., Kensing, F., 1998. Make Room for Ethnography in Design!: Overlooked Collaborative and Educational Prospects. SIGDOC Asterisk J Comput Doc 22, 20–30. doi:10.1145/571773.571781