**UNIVERSITY OF OSLO**
**Department of informatics**

# Design of Genetic Classification Software: The Case of Representation of Research References

## Master thesis

60 credits

**Jorun Børsting**

**27. November 2014**

2014

Design of Genetic Classification Software:  The Case of Representation of Research References

Jorun Børsting

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

# Abstract

*Background:* The development of genetic classification software for laboratory engineers' and laboratory doctors to aid in the diagnostic classification of patient genetic tests for hereditary diseases. The software needs to be robust enough to handle the new sequence technology (HTS) that increases users' workload, the need for handling complexity and the need for informatics systems to support their work. Lab engineers and lab doctors use both complex domain knowledge and tacit knowledge in their practice. For the clinical genetic variant classification software to support their work successfully, they need to be involved in the development process.

*Objective:* This study examines the development of genetic classification software identifying user needs and improving some of the functionalities of the software. The first phase of the research was focused on getting to know the design context and identifying the task users experienced to be most challenging. This phase identified evaluation of research references that was used in the classification of gene variant mutations as an important challenge. The second phase of research, then, focused on how the research evaluation functionality of the software should be designed to support the users work best.

*Methodology:* User-centered design with user participation, where users have high professional and domain knowledge, was deployed to involve users in the design process.

*Results:* The findings in this thesis indicate that the combination of the approaches used may be a good way of overcoming usability challenges when working in complex domains. User-centered design in combination with actor-network theory, design theory and emerging knowledge processes get a deeper understanding of the situated design context. 24 design issues and themes for diagnostic genetic research article evaluation was identified through user studies. The main findings are the multiple user strategies for deciding which article from a list references to evaluate first. One strategy deployed by the users is to manually search the research articles content with the name of the gene variant targeted in the classification. The strategy is used to investigate if the variant is mentioned in the article and how many times.

*Conclusion:* This study indicates that the article evaluation functionality needs to support diverse users' strategies deployed to evaluate research articles, to create an interrelationship between the technology and their work practice. The research gives a unique contribution to HCI research by displaying how user-centered design is both applicable and beneficial for systems in the complex domain of clinical genetic variant classification.

# Acknowledgements

There are many people who contributed to this thesis in various ways. First I would like to show my gratitude to my supervisor Alma Leora Culén, I feel I could not have done this without you. Being the only interaction design student working with this project, the discussions and guidance provided by you were highly cherished. Thank you for your understanding and help along the way.

I would also like to thank my supervisor Margunn Aanestad for your inputs and feedback.

A big thanks to Thomas Grünfeld and Morten Christoph Eike for explaining me the necessary domain knowledge of genetics. Thanks for your time and patience.

Thanks to the wonderful lab engineers and lab doctors that participated in the user studies. Thank you for your valuable time and opinions. Your contribution to this thesis is greatly appreciated.

Finally, I would like to thank my family, especially my parents and grandparents for your continuous support. Thank you for always believing in me.

Thanks to my beloved daughter Evelinn for having enriched my life with your beautiful reflections and sense of humor.

Oslo, 27. November 2014

*Jorun Børsting*

# Table of contents

# List of figures

# List of tables

# 1 Introduction

The use of genetic testing based on DNA sequence data for diagnostic and prognostic purposes is already adopted into clinical practice both internationally and in Norway. The laboratory engineers have to manually annotate the raw sequence data supported by many different informatics tools. The current clinical practice of genetic testing is just the beginning, since there has been developed a new technology (HTS) that increases the raw sequence data to massive amounts. This increases the workload and the need for handling complexity for the users and the demand for informatics systems to support their work. The use of HCI-methods and research to aid in the development of such systems is crucial to avoid usability problems and developing efficient systems. The laboratory engineers working with the genetic testing are users with high levels of expertise; they possess both complex domain knowledge and tactic knowledge. For the design of the clinical genetic variant classification software to be successful the users have to be involved in the development process. Some of the laboratory engineers' decisions in the analysis of the sequence data are based on basic knowledge of biology others are based on knowledge from scientific research. Both of these knowledge types are emerging and changing the knowledge base that the laboratory engineer uses in their practice. Research and design theory from knowledge management systems like EKP can be used to meet the demands of the emergent knowledge base. The research in this thesis is performed by employing the design theory EKP and the actor-network theory framework to examine the research case.

## 1.1 Motivation

In the presentation of the available master thesis projects, they all seemed so exciting and interesting in their own way.

One case stood out; I was intrigued by the possibility to work with the design of a system for lab doctors and lab engineers that perform an invaluable job for people with cancer. I found it meaningful to devote my thesis research to try in some small way to contribute to making their work supported better. Involving them in the design of the systems so they can have a say on how the system will affect their work environment.

I also saw it as a chance for me to learn and improve my skills as an interaction designer.

## 1.2 Thesis Project Context

The research in this thesis is carried out in relation to the project "Norwegian clinical genetic Analysis Platform" (genAP). The project is based on collaboration between Oslo University Hospital and the University of Oslo. The aim is to establish an infrastructure to increase the use of personalized medicine in a clinical context. The infrastructure shall accommodate users' requirements for functionality, legal requirements for privacy, patient rights and data security. The project was started in the autumn of 2012 and is scheduled for completion in 2015. The project is funded by the Norwegian Research Council. The development of a clinical genetic variant classification software called genAP workbench is part of the project. At the beginning of the research in this thesis a prototype of the genAP workbench was already developed.

### 1.2.1 GenAP workbench

GenAP workbench is a system developed to support the genetic variant classification in clinical diagnostics. The workbench has functionalities that are rules-based and offer a semi-automated analysis of sequence variation data. The workbench integrates relevant external and internal digital resources and software's. The information in the workbench prototype is workflow customized and sectioned into various tabs. To limit the information available at each step of the workflow to prevent cognitive overload from presenting too much information at once. The workbench contains functionalities for frequency data, mutation databases and prediction tools for missense variants and tools for splice sites. Functionality is provided for the research literature evaluation and includes a literature evaluation module. The research articles are classified by the user in terms of quality and conclusions. The choices and information gathered through the various tabs results in an final report tab, where the user is presented with suggested classifications for each variant that are found in the genetic test that is being analyzed. After the report is approved by the user the report with the results are transferred into the in-house database

The genAP workbench was developed by a domain expert with user participation in the process. It is especially developed for the laboratory engineers and lab doctors working with genetic diagnostic testing at a research and teaching hospital in Norway. The workbench might in later stages broaden its user group to include other hospitals. There were workflow Standard Operating Procedure (SOP) documents developed. Based on current work practice of laboratory engineer and lab doctors. The SOPs' were used in the development of the workbench in addition to recurring meetings with the users and through semi-structured interviews throughout the development process of the prototype. Many of the design solutions are based on user needs, but some are also based on constraints from new technology and stakeholders needs. This iterative user involvement is equivalent to user-centered design approach.

## 1.3 Research Questions

The main research questions:

1. How can the use of user-centered design and user participation, considering EKP (emerging knowledge processes), design theory and actor-network theory framework as important in the complex context of genetic analysis platform design, support the design process of clinical genetic variant classification software?

To answer this research question two sub research questions are used to get knowledge about the research case:

1. Which tasks are most challenging when using the clinical genetic variant classification software?
2. What are the important design implications for the functionality of identified task?

## 1.4 Research Methods

The research in this thesis follows a user-centered design with user participation. The research was performed in two phases. In the first phase the goal was to get to understand the research domain, and understand the intended software user group. To gain an understanding of the user there was a need to gain knowledge about the user's field of expertise, bioinformatics in this case, and current practice they engage in. Such knowledge was gained, on one hand through the literature, and on the other from users themselves, through semi-structured interviews with users, observations and conversations with domain experts. Information gathered from this phase of research is presented in Section 6.1 of this thesis. This phase was also useful for constructing a mental model of the network of actors and actants, representing tasks, software, users, policies etc. in relation to genetic testing.

In the next phase of research, I had to narrow down my research focus, again starting broadly from the proposed new PowerPoint software prototype (Eike, Skorve, Håndstad, Fontenelle, Børsting, Aanestad, Culén, Grünfeld & Undlien, 2014) and understanding users, who are expert genetic analysts, their present workflow and their understanding of the proposed software. It became clear that further narrowing down is needed, and I chose handling of research references as the main focus for the rest of my work.

# 2 Literature review

Literature searches did not identify any HCI research of clinical genetic variant classification software. There was no research of user-centered design of systems similar to the genAP workbench. Some literature concerned usability studies of developed comparable systems. It was identified an increasing awareness of the need of user involvement in development and usability testing of bioinformatics systems. Together with a growing concern that current usability approaches are inadequate for these systems.

## 2.1 Situate the research

The design of an interface for a system like the clinical genetic variant classification software in this study, have to deal with high levels of complexity, uncertainties and providing optimal decision support for domain experts. Errors or usability issues in the system have the possibility of leading to medical errors with severe consequences for patients. The development of such a system touches many interesting topics from a Human–computer interaction (HCI) - perspective (presented in table 1). These topics relate to fields of HCI and Information Systems (IS) research. From these fields, different theories and frameworks can be applied to aid in the design and analysis of the system. In addition to the use of well-established HCI design principles to identify the features and design elements that should be provided to increase the overall usability of the system.

| Relevant topics | Topics or fields of HCI and IS research |
|---|---|
| The system most support the work of the laboratory engineers and doctors. | Clinical decision support system |
| The users are domain experts. | Domain knowledge. |
| The users deal with big data. | Cognitive overload. |
| Interaction with data that users engage with in complex situations. | Analytical reasoning<br>Decision making<br>Visualization<br>Interpreting<br>Sense making<br>Planning |
| The user makes diagnostic judgments based on the data presented in the system. | Complex problem solving<br>Knowledge Discovery in Data |
| Through the use of the system a database of knowledge is built and has to be managed. | KMS /Emergent knowledge processes |
| The process of annotating the variants is done in collaboration with other clinical experts. | Groupware<br>Communication |

**Table 1: The identified relevant topics and fields of research.**

Based on the topics and fields identified in table 1, I position my research in figure 1 in relation to HCI and IS research.

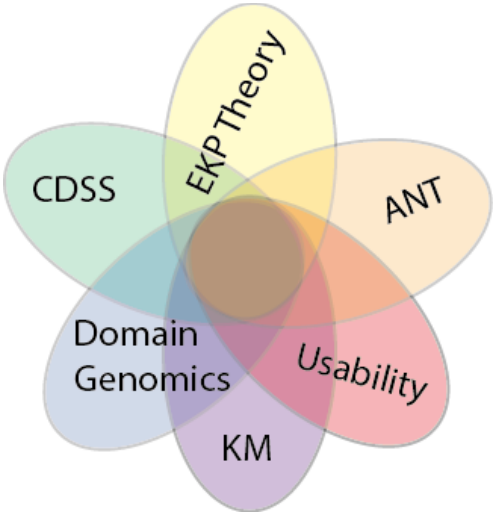



**Figure 1: Shows the placement of my research in the middle and how it relates to relevant and larger fields of study.**

I had expertise in usability and some experience in design from theory. The other three I had to study. Literature from these fields was reviewed as a preparation. To better understand how these fields of study relate to my research, I will give a short presentation of relevant research in these fields in this chapter.

# 2.2 Domain GENOMICS

The use of genetic testing for gene mutations that gives a patient an increased risk of developing cancer is now widely applied in clinical practice (Plon et al., 2008, p. 1282). There has been a new technology developed called next generation sequencing (NGS) or high throughput sequencing (HTS) technology. The introduction of HTS technology into clinical practice entails many possibilities for improving patients' health, but it is not without its challenges. The health benefits for patients lie in the improved diagnostic classification, prognostication, and the therapy selection for many diseases. Some of the challenges for integrating HTS technology into clinical practice are the lack of bioinformatics personnel, legal issues, ethical considerations and educational challenges. Another challenge is how to report the results "in a manner that is accessible and readily useable by laboratory engineer and health care professionals" (Kassahn, Scott, & Caramins, 2014, p. 414). Boyd state that the "transition of new DNA sequencing methodologies to the clinical laboratory is under way and is likely to have a major impact on all areas of medicine" (Boyd, 2013, p. 383). One of the challenges is the enormous amount of data produced by the technology. This is often referred to as the "Big data" problem. The cost of human genome sequencing can now be obtained for

less than US$5,000. This price is somewhat misleading since the process of analyzing the data is the most time-consuming and challenging. So "at present, our ability to sequence DNA by far exceeds our ability to analyze and apply the results to clinical care" (Kassahn et al., 2014, p. 413). Mardis emphasizes this with the concept of the $1,000 genome and the $100,000 genome analysis (Mardis, 2010). To obtain the patients human genome data is "only the starting point for an involved process of data analysis and further validation, particularly if clinically important decisions will be based on the sequence data" (Boyd, 2013, p. 383).

The gene variants found through genetic testing are either clearly pathogenic, clearly neutral, or variants of unclear clinical significance (VUS). Kassahn et al. state that "a typical whole-exome sequence returns thousands of variants of unknown clinical significance" (Kassahn et al., 2014, p. 414). The VUS variants "present a considerable challenge to the diagnostic laboratory and the receiving clinician in terms of interpretation and clear presentation of the implications of the result to the patient" (Plon et al., 2008, p. 1282). The laboratory engineer often manually annotates and makes clinical interpretations of the genomic variants with the support from different software tools. In the annotation process the laboratory engineers makes classifications based on knowledge from biology (like the effect of a given variant on protein function) and based on scientific research of the variant. That is, if there exists research on the specific gene variant and if the research is good enough. Kassahn et al. state that "There is currently limited evidence and agreement on which variants are clinically actionable and medically relevant" (Kassahn et al., 2014, p. 414; Manolio et al., 2013).

Boyd state that "one of the most important questions is how to deal with rare sequence variants (…) that may be important for an individual patient's disease pathogenesis, prognosis, or therapeutic response, but whose significance is poorly understood because it has not been observed in enough people or studied under the conditions of a controlled clinical trial" (Boyd, 2013, p. 388). He further stated that "these variants will undoubtedly pose a significant challenge for the responsible interpretation of genome-sequence information in clinical settings" (Boyd, 2013, p. 388). The amount of scientific research about gene variants is growing. In the future we will know much more about the genetic variants underlying human disease. Boyd argues that "It also seems likely that large numbers of sequence variants will resist interpretation, and remain a kind of "background noise" of no definite significance in each patient's genomic record" (Boyd, 2013, p. 387).

Kassahn et al.state that "For the first time, clinical genetics and genomics is not limited by our ability to sequence, but our ability to clinically interpret and use genomic information in health management" (Kassahn et al., 2014, p. 413). He also argues that "the development of standardized variant annotation and interpretation approaches and software tools implementing these warrants further support" (Kassahn et al., 2014, p. 413). Shyr et al. state that "This transition is challenging; identification of potentially causal mutation(s) amongst $\sim 10^6$ variants requires specialized computation in combination with expert assessment" (Shyr, Kushniruk, & Wasserman, 2014, p. 1). Others also highlight the need for informatics systems to aid the annotation and clinical interpretation of genomic variants (Collins et al., 1998; Khoury et al., 2007; Sim et al., 2001; Stanley, Sunyaev, Greenblatt, & Oetting, 2014;

Torkamani, Scott-Van Zeeland, Topol, & Schork, 2011; Welch, Loya, Eilbeck, & Kawamoto, 2014).

## 2.3 Knowledge management

Through the process of annotating gene variants in the system, there is built a database containing a body of knowledge of gene variant classifications. This knowledge is constantly growing as new variants are found and classified. Boyd states «although this body of knowledge is already becoming complex and detailed, the remarkable discoveries so far obtained through HTS methods in almost all areas of medicine probably represent a mere glimpse of what lies ahead" (Boyd, 2013, p. 404). The knowledge is not only growing, but also changing as new scientific research emerges. Neri et al. argue that "as evidenced-based use of genetic data becomes more common for a variety of purposes, there is a growing need for information technology (IT) to support the management of the rapidly expanding knowledge" (Hamburg & Collins, 2010; Neri et al., 2012; Scheuner et al., 2009).

The importance of managing knowledge is present in many different professions. Research institutes, management, businesses administrations and libraries to name a few. The field of knowledge management has its roots from organizational theory, but has been introduced into many other fields. We live in an information rich society and the need for managing big bodies of knowledge is highly current. According to Davies et al. KMS are "the tools, techniques and processes for the most effective management of an organization's intellectual assets" (Davies, Harmelen, & Fensel, 2002, p. 2). There are many different definitions of KM. The lack of a unified definition of KM might be related to the vagueness of the concept of knowledge itself. Plato defined knowledge as "justified true belief"(Kakabadse, Kakabadse, & Kouzmin, 2003). Many definitions of knowledge have since followed. Prusak defined knowledge as "a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms" (Davenport & Prusak, 1998, p. 5). One often cited definition of knowledge in KM is coined by Alavi and Leidner (2001) that defines KM as "a dynamic, continuous organizational phenomenon of interdependent processes with varying scopes and changing characteristics" (Alavi & Leidner, 2001, p. 131).

How knowledge is defined have impact on how knowledge management should be perceived. Based on KM literature in different fields Alavi presents an overview of different perspectives on knowledge and their implications on KM and KMS. Alavi and Leidner (2001) present three major points from the review of different knowledge perspectives:

1. A great deal of emphasis is given to understanding the difference among data, information, and knowledge and drawing implications from the difference.

2. Because knowledge is personalized, in order for an individual's or a group's knowledge to be useful for others, it must be expressed in such a manner as to be interpretable by the receivers.
3. Hoards of information are of little value; only that information which is actively processed in the mind of an individual through a process of reflection, enlightenment, or learning can be useful.

In the description of knowledge Alavi and Leidner (2001) states that "knowledge may be tacit or explicit; it can refer to an object, a cognitive state, or a capability; it may reside in individuals, groups (i.e., social systems), documents, processes, policies, physical settings, or computer repositories" (Alavi & Leidner, 2001, p. 131). He argues that since there isn't a single definition of knowledge, there is not one single optimal approach to organizational knowledge management. Rather a "variety of knowledge management approaches and systems needs to be employed in organizations to effectively deal with the diversity of knowledge types and attributes". If knowledge is viewed as a state of mind, then KM involves enhancing individual's learning and understanding through provision of information. And the "role of IT is to provide access to sources of knowledge rather than knowledge itself" (Alavi & Leidner, 2001, p. 111). From the perspective of Knowledge as vis-a-vis data and information. Data is facts and raw numbers. Information is processed and interpreted data and knowledge is personalized information. Then KM and IT focus on exposing individuals to potentially useful information and facilitating assimilation of information (Alavi & Leidner, 2001, p. 111).

If knowledge is viewed as an object to be stored and manipulated, "then knowledge management should focus on building and managing knowledge stocks" (Alavi & Leidner, 2001, p. 111). If knowledge is viewed as a process of applying expertise, KM focus is on knowledge flows and the process of creation, sharing, and distributing knowledge. Role of IT is to provide link among sources of knowledge to create wider breadth and depth of knowledge flows (Alavi & Leidner, 2001, p. 111). The view of knowledge as a capability, "where knowledge is the potential to influence action. Suggests a knowledge management perspective that is centered on building core competencies, understanding the strategic advantage of know-how, and creating intellectual capital" (Alavi & Leidner, 2001, p. 111).

The four basic processes in knowledge management is creating, storing/retrieving, transferring, and applying knowledge. These processes are distinct but interdependent. The user of KMS can simultaneously be involved in multiple knowledge management process chains. Therefore Alavi argue that "knowledge management is not a monolithic but a dynamic and continuous organizational phenomenon" (Alavi & Leidner, 2001, p. 131). Due to "Knowledge management's complexity, the appropriate approach(s) of knowledge management, resource requirements and software tools are reflected in the characteristics of the specific knowledge management process" (Alavi & Leidner, 2001, p. 131).

## 2.4 Clinical decision support tools

Clinical decision support systems (CDSS) provide clinicians with tools to enhance decision-making in their clinical workflow. The systems rely on intelligently filtering to provide the right information at the appropriate time. The aim of CDSS is to support and improve the clinicians work practice in order to enhance health and health care. These tools can include diagnostic support, clinical guidelines and computerized alerts and reminders to health care personnel and patients. Different types of tools and CDSS are developed for different types of health care and work practices.

Kawamoto et al. defines a CDSS "as any electronic or non-electronic system designed to aid directly in clinical decision making, in which characteristics of individual patients are used to generate patient-specific assessments or recommendations that are then presented to clinicians for consideration" (Kawamoto, Houlihan, Balas, & Lobach, 2005, p. 1). They include both electronic and non-electronic systems in their definition. Kawamoto et al. performed a literature analysis based on research articles from 1990 to 2004, examining the features that correlated with the decision support systems' ability to improve patient care. Seventy studies were included, and the research indicated that "decision support systems significantly improved clinical practice in 68% of trials" (Kawamoto et al., 2005, p. 1).

Kawamoto et al. identified 15 potentially important features of CDSS, and then they performed a literature analysis of 70 articles of research on CDSS. They identified five features were "the success rate of interventions possessing the feature was significantly greater than that of interventions lacking the feature" (Kawamoto et al., 2005, p. 3).

Decision support should be provided automatically to clinicians. The importance of integrating the system into the users' workflow. It was also found that the system should record a reason from the clinician when they did not follow the advised course of action. Their findings argue for developing systems that not only provides an assessment of the patient, but provide a recommendation for how the health issue should be managed (Kawamoto et al., 2005, p. 3)

Lindgren argue that "developing CDSS for health and medical care is essentially a process of developing the clinical practice in which the system will be used. This means that there are several stakeholders that need to become involved who have influence on the results" (Lindgren, 2011, p. 129).

Lindgren also points out "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 use environment is assessed, and requirements for a CDSS are formulated" (Lindgren, 2011, p. 129).

## 2.5 Actor-network theory

There are different views on ANT and its application. ANT is described as it is used in this thesis.

Actor-network theory (ANT) is a theoretical framework from the social sciences. ANT was developed through the work of several researchers, including Bruno Latour, Michel Callon and John Law. ANT is used to analyze and explore relationships which are simultaneously material (between things) and semiotic (between concepts), to better reveal the

complexities of our sociotechnical world (Cressman, 2009, p. 2; Law, 2009). All the social-technical elements are included in networks of actors/ actants. In this thesis the word actor is used. It is best known for the controversial issue of including the non-human relationships as part of the social as a quest for symmetry, where human and non-humans are integrated into the network (Aanestad, 2003, p. 7). The social-technical focus is described by Latour (1990) "Contrary to the claims of those who want to hold either the state of technology or that of society constant, it is possible to consider a path of an innovation in which all the actors coevolve" (Latour, 1990, p. 117). Through the ANT analyze concepts that are perceived to be social might be revealed as partly technical and the technical might be partly social (Tatnall & Gilding, 2005, p. 957). ANT therefore "deny that neither purely technical or purely social relations are possible" (Tatnall & Gilding, 2005, p. 957). This draws attention to analyzing both the visible (the technical/objects) and the invisible (the social) and identifying the relationships between these. These relationships can be both material and semiotic simultaneously and combined into a network of actors that acts as a whole (Law, 2009). In ANT the technology and other actors gets their role as they are put in use and integrated in the network. (Aanestad, 2003) If the technology is not used it loses its role in the network. The network is seen as heterogeneous and consists of actors of different materials both human and non-human. Tatnall and Gilding (2005) argue that "An actor-network is configured (Grint & Woolgar, 2013) by the enrolment of both human and non-human allies, and this is done by means of a series of negotiations in a process of redefinition in which one set of actors seek to impose definitions of the situation on others" (Gallon, 1999; Tatnall & Gilding, 2005, p. 962).

Tatnall and Gilding (2005) argue that ANT can be particularly useful for studies in areas that involve a consideration of some of the social and political issues in information systems "like interface design, usability testing, the use of distributed systems within organizations" (Tatnall & Gilding, 2005, p. 963). By exploring the nature of the technology that is developed and implemented in relation to the other relations involved, the researcher might gain a deeper and broader understanding of the design process.

## 2.6 EKP design theory

Theory is used to aid the design and analysis process in human computer interaction (HCI). Design theories "make the design process more tractable for developers by focusing their attention and restricting their options, thereby improving development outcomes" (Markus,

Majchrzak, & Gasser, 2002, p. 180). Another advantage is that IS design theories inform researchers by suggesting testable research hypotheses" (Markus et al., 2002, p. 180). There is different design theories related to specific technology. An example is DSS design theory that "makes the DSS design problem more manageable for developers, and it gives researchers a basis for making predictions about DSS use patterns and impacts" (Markus et al., 2002, p. 180). Design theory for decision support systems "tells developers not to specify decision-making problems as procedural processes and not to try to specify all user requirements in advance of starting system development" (Markus et al., 2002, p. 180).

Emerging knowledge processes (EKPs) are organizational activity patterns that exhibit three characteristics in combination: an emergent process of deliberations with no best structure or sequence; requirements for knowledge that are complex (both general and situational), distributed across people, and evolving dynamically; and an actor set that is unpredictable in terms of job roles or prior knowledge. Examples of EKPs include basic research, new product development, strategic business planning, and organization design.

(Markus et al., 2002, p. 180) developed an IS design theory for emerging knowledge processes support systems called EKP design theory. Their theory present requirements for IT Support of EKPs based on characteristics of emergent knowledge processes.

In emergent knowledge processes it is nearly impossible to predict in advance who will participate in the process and which tools they will use. Therefore systems cannot target specific user roles, depend on training, or assume motivation to use the tool. The EKP design theory argues that since knowledge in emerging knowledge processes support systems is distributed and includes both general expertise and local context knowledge. Then systems must accommodate complex, distributed, and evolving knowledge-bases. Since the process is emergent, "the systems must support an unstructurable, dynamically changing process of deliberations and tradeoffs" (Markus et al., 2002, p. 206).

Markus et al. (2002) developed 6 principles to aid in the design and development of EKP Support Systems:

1. Design for customer engagement by seeking out naive users.

2. Design for knowledge translation through radical iteration with functional prototypes.

3. Design for offline action.

4. Integrate expert knowledge with local knowledge sharing.

5. Design for implicit guidance through a dialectical development process.

6. Componentize everything, including the knowledgebase.

(Markus et al., 2002, p. 206)

Markus et al. argue that the “knowledge-intensive emergent processes have challenging information requirements” (Markus et al., 2002, p. 184). The requirements are quite different from those of semi-structured business processes. Firstly since documents that are used in emergent processes to search for the information they need are often poorly indexed and stored in contrast to analyzing numeric data presented in tables and graphs. Secondly since much of the knowledge involved in sense-making processes are tacit, not explicit. This makes the knowledge much more difficult to capture and share. Thirdly since “knowledge-intensive emergent processes have a high level of expert knowledge content. “and “when tacit knowledge can be made explicit, it cannot easily be represented numerically, but must instead be represented as if-then rules (Baligh, Burton, & Obel, 1996) , as cases (El Sawy & Bowles, 1997) , or as text.” It is further stated that “because non-users may not understand expert jargon, the knowledgebase must be translated into terms non-experts can understand” (Markus et al., 2002, p. 183). Markus et al. also argues that” in most knowledge-intensive emergent processes, knowledge is distributed cross many different people (Hutchin, 1991). Some of the distributed knowledge is local (e.g., conditions in a certain geographic locale), and some is general (e.g., scientific knowledge). Unless distributed expertise can effectively be brought together during what Hutchins calls "local design activities," knowledge will be incomplete, and action faulty” (Markus et al., 2002, p. 183).

Markus et al. states that “expert systems do include general expert knowledge, but they may not support contextualization, and they often sacrifice the flexibility needed for process emergence” (Markus et al., 2002, p. 184). While knowledge-intensive emergent processes “require knowledge and expertise in applying the knowledge. They require tacit and explicit knowledge, general and contextual knowledge. Because knowledge is distributed, they require knowledge sharing” (Markus et al., 2002, p. 184).

## 2.7 Design of bioinformatics systems

In the development of bioinformatics systems, “there is evidence of an increasing awareness of the need of usability studies in the development of biomedical systems in general and of the benefits that a systematic user-centered design process can bring to the development of interactive systems in the bio-related areas” (Bolchini, Finkelstein, Perrone, & Nagl, 2009, p. 654).

Bolchini et al. (2009) carried out usability evaluations of existing bioinformatics systems, with the aim of identifying usability problems for bioinformatics systems and propose design solutions for bioinformatics systems in general. They state that the contributions made by others have focused on important aspects of improving the user experience of biological databases, but argue that “little has been done to analyze the underlying design characteristics of web bioinformatics resources that can lead to potential usability problems”. They further state that “tackling design issues identify the usability problems at their source, and helps to prevent the emergence of problems in current and future applications” (Bolchini et al., 2009, p. 407).

They performed a usability inspection of the bioinformatics website "CATH—Protein Structure Classification" focusing on identifying usability issues related to navigation and information architecture designs. Secondly they performed user testing on BioCarta, NCBI and Swiss Prot focusing on search task usability issues. The focus was on identifying overall usability problems of bioinformatics system and not target at the individual website. They identified the categories "search scope", "search Ontology" and "query syntax", as usability concerns that needed to be addressed related to the users search tasks. Stating that "When interpreting search results, the main problem encountered by the users was the difficulty in managing long lists of results" (Bolchini et al., 2009, p. 411). They divided this issue into three sub issues. Firstly the "visual organization of the result items makes the search results appear confusing to users, and does not guide their eyes to easily master the complexity of the results at a glance." Secondly the ranking criteria for the list of search results were not clear. "They hope the order is based on relevance, but how this relevance is obtained is not communicated to them and they do not find the elements to understand it" (Bolchini et al., 2009, p. 411).

They argue that it is important "to address this problem, it is important to explicitly communicate to the user the actual ranking criteria used for displaying the results (as results are displayed) and, possibly, to allow sorting the obtained results by multiple, additional attributes (e.g. by publication/release date, by alphabetical order)."

Thirdly they state that" all users found severe obstacles in understanding the actual content of the documents (before clicking on it) on the basis of the representative information appearing in the list (e.g. the document title or the document body excerpt did not help much in identifying relevant content). In some cases, users found that specific symbols were associated with each list item and could not find any explanation for them. For example, letters 'H' and 'M' are displayed aside each BioCarta search result without any legend or textual clue" (Bolchini et al., 2009, p. 412).

They argue that they "have identified and characterized a sample of usability issues potentially relevant to web bioinformatics resources, in general. These specifically concern the design of the navigation and search mechanisms available to the user" (Bolchini et al., 2009, p. 412).

Bolchini et al. (2009) state that the "usability of web bioinformatics applications can be significantly improved by leveraging proven practices in usability and web engineering" and that their "ultimate goal is to make available proven design patterns(i.e. proven solutions that work) and conceptual tools in order to promote a more aware human-centered development process of bioinformatics applications" (Bolchini et al., 2009, p. 412).

Neri et al. conducted a usability evaluation of an application called GeneInsight Clinic, which was designed "to assist providers in receiving and managing a patient's genetic profile, including ongoing updated interpretations of the genetic variants in those patients" (Neri et al., 2012, p. 950). Managing the knowledge of gene variant classifications that are constantly evolving is challenging. Neri et al. (2012) "conducted usability tests with potential users of

this application and reported findings to the application development team, many of which were addressed in subsequent versions" (Neri et al., 2012, p. 950).

The usability tests revealed some issues related to the interface of the system. Neri et al. state that "Almost all observed usability issues fell under the following categories: icon inconsistencies, labeling and language issues, and placement and organization." They argue that since the issues required relatively little development effort to fix, "the consistently incorporating this type of analysis in the development process can be highly beneficial" (Neri et al., 2012, p. 950). And state that "Our results demonstrate that using a development and design process that is user focused helped optimize the value of this application for personalized medicine" (Neri et al., 2012, p. 950).

## 2.8 Usability study of clinical next-generation sequence analysis software

Shyr et al. (2014) performed a study, which they stated was "the first application of usability methods to evaluate software interfaces in the context of exome analysis" (Shyr et al., 2014). The methods they used where surveys, interviews and cognitive task analysis. The usability study was performed on two user interfaces for clinical next-generation exome sequence analysis software named Varsifter and KGGSeq. The subjects in the study consisted of 10 clinical geneticists that interacted with the software using scenario cases and the "think aloud" method. The sessions were recorded and performed at the subjects' workplace at a research and teaching hospital. The objectives of the study were twofold. Firstly to "ascertain the key features of successful user interfaces for clinical exome analysis software based on the perspective of expert clinical geneticists." Secondly to "assess user-system interactions in order to reveal strengths and weaknesses of existing software, inform future design, and accelerate the clinical uptake of exome analysis." They further wanted to explore "the impacts of different user interface designs on analysis workflows and outcomes" (Shyr et al., 2014, p. 130). Their study detected 193 usability issues the majority of which concern interface layout and navigation, and the resolution of reports. They divided the detected usability problems into 5 main themes. The themes where further divided into 12 sub categories. The main themes where visualization, information, system response, functionalities and overall usability. The theme visualization had the sub categories navigation, layout, operation consistency and graphics. For Varsifter "every clinician complained that text or functions were hidden from view due to scrollbars and/or hidden panels" (Shyr et al., 2014, p. 132). The Information theme had the sub categories resolution, label and system messages. "For KGGSeq, there were multiple complaints regarding the use of bioinformatics jargons that the clinicians were not able to comprehend" (Shyr et al., 2014, p. 133). System response had the sub categories response time and system status. The software should provide an approximate processing time whenever possible and a clear indication of system status. Shyr et al. argue "as the community moves to whole genome data, the resulting size and complexity will exacerbate concerns about the speed of processing – thus it is critical for the software to

provide time estimates whenever a job cannot be completed rapidly (i.e. >10 s)" (Shyr et al., 2014, p. 135). The theme functionalities had the sub categories compatibility and scope of functionalities. Shyr et al. (2014) states that it was identified that a majority of functionality problems were related to the clinician's inability to execute the software's implemented function. Overall usability had the sub categories overall usage and total. The frequency of errors was much higher for the systems with no ability to revert from one chosen state to a previous state without having to restart the entire analysis.

Shyr et al. (2014) state that their study highlights gaps in specific software features typical within exome analysis. Based on their research they present the following implications and desired features for the design of clinical exome interpretation software:

1. Rich filter functionalities (i.e. variant calls with simple column-based filtering are insufficient)
2. Software design structured with focus on genetic models (e.g. Mendelian inheritance)
3. User defined workflow management with stepwise reports
4. Fast response time with estimates given for wait steps
5. Team support to allow multiple clinicians to annotate/ review data
6. Interoperability with widely used online resources/databases and data formats
7. Frequent updating to support emerging tools, data standards and input types.

They call these 7 implications the clinical exome interpretation software user desiderata. That "represent a key feature set for future systems to deliver." (Shyr et al., 2014, p. 133)

Shyr et al. (2014) argue that they "find that a major impediment for adoption of exome analysis software is a lack of clear presentation (organization), description and help messages for the provided functionalities. Non-computational healthcare professionals will not choose to adopt a software package unless the functionalities are easily executable and can fit into a clinician's workflow." And states further "recurring domain-specific usability challenges need to be addressed" (Shyr et al., 2014, p. 133).

### 2.8.1 Reported user recommendations and requests for key functionality

Shyr et al. states that "a key observation from the study is the importance of supporting diverse workflows for the range of potential genetic hypotheses. Specifically, the system should be structured around the commonly used analysis models, such as Mendelian recessive inheritance" (Shyr et al., 2014, p. 129). Clinicians value such structured approaches, as they are expected to follow standardized protocols in their practice" They also state that "the clinicians perform best when the flow of the system is structured into well-defined yet customizable layers for incorporation within the clinical workflow" (Shyr et al., 2014, p. 129). But warn 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. 129). The software should give the users possibility to define their own custom workflow, providing step-wise reports so the impact of each step can be assessed.

### 2.8.2 Present variants in a tabular format but retain flexibility in layout

Shyr et al. argues that "displaying the variations visually in a tabulated form with sortable columns allows the clinician users to browse and prioritize the data." Pointing out that "another advantage of tabular structure is it is highly similar to Excel representation, a program that is frequently used by clinicians" (Shyr et al., 2014, p. 134). Some of the clinicians also stated that they would like the order of the columns to be adjustable so that they can customize the type and order of information presented.

### 2.8.3 Allow customizable filtering pipelines and prioritizing strategies

It is important in exome and genome analysis with filtering and variant prioritization. Shyr et al. argue that "the software should provide an intermediate output to evaluate the effectiveness of a particular filtering step, and the ability to return to the previous result or continue to the next depending on the context of that intermediate output" (Shyr et al., 2014, p. 134). By providing this iterative design feature, the amount of slips is reduced, and the user can investigate the data under different scenarios.

### 2.8.4 Support collaborations and team-based communications

Each case is evaluated by more than one specialist, so it is desired for clinical exome interpretation software to support team collaborations for collective annotation and review of data. Shyr et al. states 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).

Shyr et al. (2014) found that most exome analysis software, both free and commercial, do not provide suitable functionalities for facilitating multiple users to collaborate on the same data. Shyr et al. state "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" (Shyr et al., 2014, p. 134). They further argue that software that empowers collaborative analysis would be well received.

### 2.8.5 Compatible with multiple data formats

Shyr et al. state that “being interoperable with the data standards and currency with updates is important for widespread adoption” (Shyr et al., 2014, p. 134). The task of accessing all the external online resources in the process of annotating the variants was reported to be amongst the most time-consuming steps in their work practice. Thus, they state ”the capacity of software to automate data mining of these resources may accelerate analysis and increase success rates” (Shyr et al., 2014, p. 134).

### 2.8.6 The use of user studies

They argue the "results highlight how the study of user responses can lead to identification of usability issues and challenges and reveal software reengineering opportunities for improving clinical next-generation sequencing analysis” (Shyr et al., 2014, p. 129).

Shyr et al. (2014) argue "the results are general and should inform active and future software development for genome analysis software. “The ultimate objective of their study was to use the two software as a starting point to collect and understand general exome analysis interface characteristics. These could aid in the development of the next generation of clinical genetics exome analysis tools. They also argue that their results indicate which features and requirements the community desires in future software. The further argument is that “the results highlight opportunities to dramatically accelerate clinician analysis and interpretation of patient genomic data”(Shyr et al., 2014, p. 129)

## 2.9 Usability practices in complex domains

There is a growing concern that current usability approaches are inadequate for evaluating complex, domain-specific tools. This is illustrated by the statement “Consider a project for evaluating how consumers shop using an e-commerce website (“everyday domain”) versus scientists using a bioinformatics analysis tool (“complex domain”). How does a usability professional trained as a generalist deal with these seemingly two different projects?” (Chilana, Wobbrock, & Ko, 2010, p. 2337)

Chilana et al. (2010) conducted semi-structured interviews with 21 usability professionals working with medical imaging, software development, network security, aviation, healthcare, test and measurement devices, genomic analysis, financial derivatives, statistical analysis, and business-process support.

Chilana et al. “investigated how empirical usability testing in these domains differs from conventional GUI and web applications, and how it changes the dynamics of collaborations among usability professionals, users, and developers” (Chilana et al., 2010, p. 2338). Their aim is to establish a greater understanding of usability evaluation in complex domains and reveal the challenges commonly posed by domain complexity and how practitioners work

around them. Chilana et al. recorded the sessions and analyzed the data, by following “an iterative process of applying open coding and axial coding to discover relationships among emerging concepts in the data, followed by selective coding to integrate the results” (Chilana et al., 2010, p. 2337).

Chilana et al. (2010) found that despite the best efforts by usability professionals to get familiar with complex domains on their own, the lack of formal domain expertise can be a significant hurdle for carrying out effective usability evaluations. Partnerships with domain experts led to effective results when domain experts were willing to be an integral part of the usability team. Chilana et al. argue that these "findings suggest that for achieving usability in complex domains, some fundamental educational changes may be needed in the training of usability professionals" (Chilana et al., 2010, p. 2337).

Chilana et al. (2010) states that Human-Computer Interaction (HCI) researchers, practitioners, and educators hold opposing opinions about usability in complex domains. Some argue that usability work in any domain requires a learning curve so complex domains are not fundamentally different. Others state that complex domains are different, but the issue is resolved with ethnography and HCI field methods for observing work, eliciting requirements, and understanding domains. Chilana et al. (2010) state that usability practice in complex domains are often methods like usability testing and heuristic evaluations.

Chilana et al. (2010) presents three key insights about usability evaluation in complex domains:

1. Even highly experienced usability professionals regarded work in complex domains to be more challenging than working with GUIs and web applications. Some went to extreme measures to understand a complex domain—even taking night classes—but generally found their efforts to be insufficient.
2. To compensate for their lack of domain expertise, some usability professionals tried to learn from domain experts upfront and in depth. Some developed partnerships in which domain experts acted as consultants, contributing on an infrequent but regular basis. Others formed deeper, persistent relationships, incorporating domain experts into their team and working hand-in-hand through all aspects of their evaluations.
3. In addition to challenges in gaining credibility with developers, usability professionals designing for complex domains had the additional challenge of gaining credibility in their domain expertise. Some managers admitted that they were becoming reluctant to hire usability experts who lacked significant domain expertise.

Based on these key insights Chilana et al. (2010) argues that “domain complexity clearly imposes new knowledge demands and the need to change dynamics of collaboration in usability practice” (Chilana et al., 2010, p. 2338).

Based on the analysis of their data Chilana et al. (2010) identified three key characteristics that are pronounced in complex domains:

1. Domain-specific terminology. There was frustration related to not understanding the domain jargon.
2. Every situation is unique. Even after several years of working in the same area, some interviewees were not able to understand all the unique situations and frequent exceptional conditions that arose in complex domains.
3. Limited access to domain experts**.** A major obstacle was getting access to the right domain experts at the right time to understand domain-related details. A related problem was finding representative users from complex domains:

Chilana et al. (2010) state that together, these three characteristics affected all aspects of usability work in complex domains.

Chilana et al. state that "all the interviewees agreed that acquiring domain expertise on their own was a good starting point, but insufficient for understanding the nuances of a complex domain" (Chilana et al., 2010, p. 2341). Hence, the interviewees expressed that they relied heavily on domain experts. The domain experts helped the interviewees to learn about and clarify domain-related details. Through data analysis, Chilana et al. (2010) identified three models of collaboration between usability professionals and domain experts. These were called iterative elicitation, persistent partnership, and upfront investment.

### 2.9.1 Iterative elicitation

Iterative elicitation was the most common form of collaboration, consisting of regular back-and-forth exchanges between usability professionals and domain experts. One challenge for usability professionals in complex domains is generating relevant questions or tasks for usability testing. Chilana et al. explained that "with iterative elicitation, the interviewees outlined tasks and then sought feedback and revisions from domain experts" (Chilana et al., 2010, p. 2343). Iterative elicitation was also used to shed light on their observations and findings. Chilana et al. argue that "simply having domain experts available for consultation was not effective in understanding the necessary intricacies to support problem-solving tasks in a complex domain" (Chilana et al., 2010, p. 2343).

### 2.9.2 Persistent partnership

Persistent partnership with domain experts was partnerships that sustained throughout the entire usability testing. From early planning to the end, they worked hand-in-hand with domain experts. Together they created and verified tasks, co-facilitate observations, and

analyzed the results. Chilana et al. state that “with persistent partnership, domain experts also co facilitated the test sessions, similar to the idea of cooperative usability testing. The domain experts not only helped interpret what was being observed, but also answered domain-related questions posed by users” (Chilana et al., 2010, p. 2343).

### 2.9.3 Upfront investment

Upfront investment was when usability professionals during the planning phase relied extensively on domain experts. The domain experts helped to understand scenarios, tasks, and the target users of the domain. After the planning phase, the usability professionals do not rely on the domain expert. This strategy was used mostly to gather design requirements or to set up usability tests later in the design process.

Chilana et al. argue that their results suggest “that either (1) usability professionals take on formal training in a particular domain, or (2) usability professionals commit to a long-term relationship within a domain, start small, and iteratively work the way up to doing more complex testing, or (3) domain experts carry out usability evaluations themselves” (Chilana et al., 2010, p. 2344).

# 3 Theoretical chapter

The research in this thesis follows a user-centered design process, and the participatory approach. Thus, users and other stakeholders have been actively engaged in the process. In this chapter, HCI design approaches are briefly described, followed by how different design approaches are distinguished based on the level of user engagement. After the brief introduction, a bit more thorough view is presented in Section 3.1

## 3.1 HCI design approaches

In HCI the understanding of the user is viewed as an important factor in systems development, which contributes to making the system better suited to users' needs and abilities and, thus, increasing chances that the system would be accepted by users. The HCI designers have a focus on identifying user needs. This is a process far more complex than simply asking the user what they want and need. Since people often don't know what that is possible. If you would ask someone in the 1950's what they would like, the chances for getting the answer the internet is very slim. If you rephrase the question and explain what the internet is and then ask if they would like it, chances are, the answer would have been different (Rogers, Sharp, & Preece, 2011, p. 334). Therefore, the designers approach in HCI is to involve the future users in the design process of systems.

### 3.1.2 The role of the user in diverse design approaches

There are different design approaches that can be used as a framework to guide, inspire and inform the researcher through the design process. The design approaches involve various levels of partnership with users during the design process. The figure 2, from Sanders and Chan (2006), illustrates how the expert mindset of user-centered design sees the user as a subject, and that the knowledge of the user is often research driven. On the opposite side is participatory design (PD), where the user is seen as an equal partner that has a critical role in the design team.

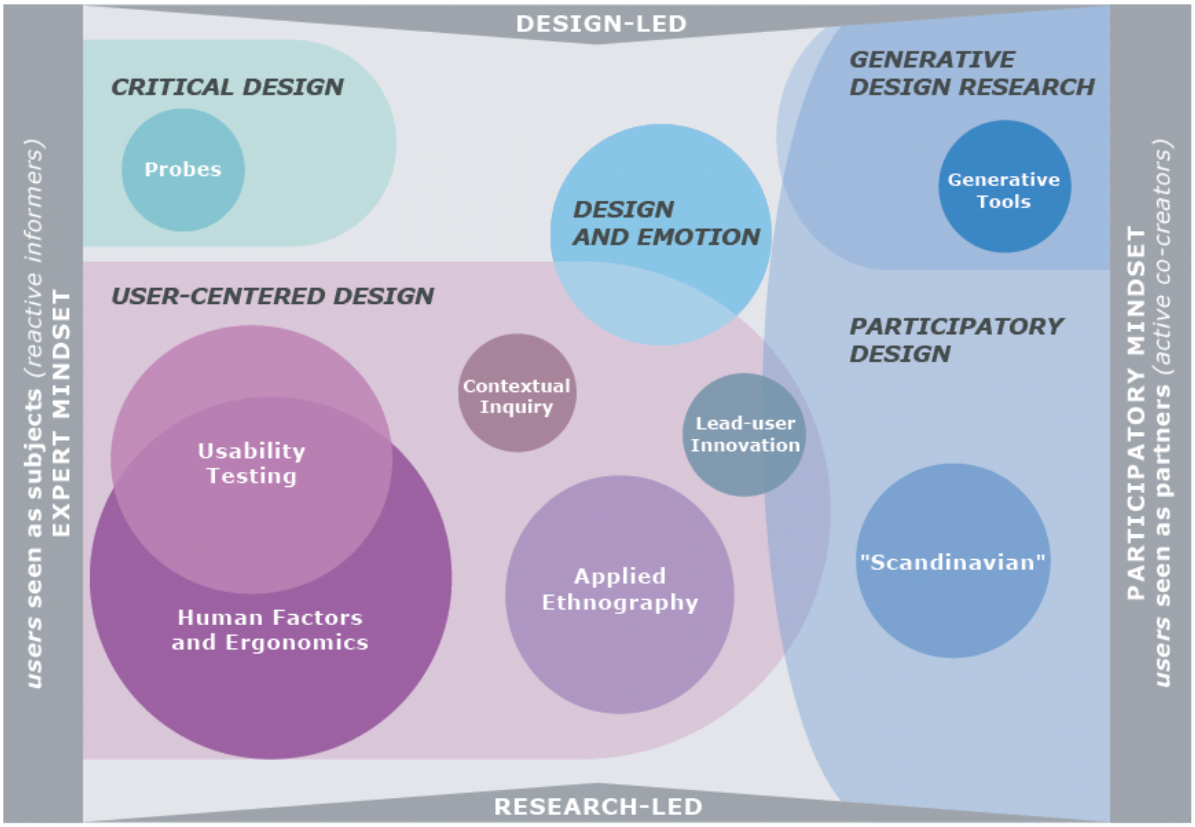



**Figure 2: Sanders and Chan (2006) presentation of the current landscape of human-centered design research.**

Sanders and Stappers state that that USD and PD have begun influencing each other (Sanders & Stappers, 2008, p. 5). The researcher can choose to use a combination of approaches as long as they are not in conflict.

### 3.1.3 User-centered design

User-centered-design (USD) is a broad term covering different design processes that are all unified in the focus that future users shall influence the shaping of a design. In USD the role of the designer is to know the users and apply that knowledge in the design process. The designer and users are interdependent, but their roles are distinct (Sanders, 2002, p. 1). The users are involved throughout the entire design process and their mental models, requirements, tasks and goals are a guiding the design process. The user is not integrated as a part of the design team, so the role of the designer is to be the voice of the users in the design process. Rogers and others emphasize the importance of the user involvement by stating the "user knows best and is the only guide to the designer; the designer's role is to translate the users' needs and goals into a design solution" (Rogers et al., 2011, p. 320). The systems are designed to support the context of use, user characteristics and user behavior. User characteristics can be general like the statistics that say that 1/12 men have some sort of color blindness. Another form of user characteristics are more related to something specific, like a certain work task. USD is about getting to know the real users and their tasks, which then is used to inform design. When the users' voices and needs are heard and taken into account in the design process. The chances for the end result being useful, meeting the users' needs and gaining user acceptance is much higher (Rogers et al., 2011, p. 328). There are multiple ways the designer can involve the users in the design process. Some examples are interviews,

observations, using performance tasks, surveys and even making the user a co-designer. The ethnography method can also be a useful source of information for a user-centered design process. There are three principles that are often used in USD presented by Gould and Lewis: 1) an early focus on users and tasks. Observing the users doing their normal tasks and understanding the nature of these tasks. Then involving the users in the design process. 2) Use empirical measurements, both of users early in the development and later in the process. 3) Use iterative design, where there is a repeatable cycle of design, test, measure and redesign. The issues found in user testing are quickly fixed and reevaluated with users to see if the issues are resolved (Gould & Lewis, 1985; Rogers et al., 2011, p. 327). User-centered design is both practical and conceptual. The design requirements are tested with the users and reevaluated throughout the entire design process. The approach is also conceptual, in that it provides a design model and user-centered design principles. The design principles frequently used are visibility, consistency, feedback, affordance and constraints.

# 4 Methodology chapter

Within research fields, different research paradigms provide the researcher with a view of the world consistent with underlying philosophical assumptions. This has led to different methodologies providing strategies of inquiry, and methods approaching how empirical material is gathered and analyzed (Finken, 2013b).

## 4.1 Choice of methodology

Both quantitative and qualitative research methods have their strengths and weaknesses. These require considerable considerations by the researcher when deciding the right research method for their research project. Quantitative research methods focus on objective data that can be statistical measured and aims at being statistically significant. Therefore, the data is collected from a large number of participants. Qualitative research methods, on the other hand, are more focused on the meanings, concepts, metaphors definitions, and descriptions. It yields more in depth research results. In quantitative research, surveys are often used. Qualitative research methods are observation, interviews, focus groups and analysis of both text and video.

Silverman (2005) argues that the right method of research should be decided based on the nature of the research question. In this thesis research, it was important to understand the domain of research and collect user opinions about their work practice and the design of the article evaluation.

## 4.2 Research paradigm

There are different research paradigms to choose from when starting a research project. A positivist research paradigm views reality as objectively given. The positivist research paradigm has an emphasis on quantitative data. Reality is described by measurable properties that are independent of the observing researcher and the research instruments. Often a theory is tested with statistical tools. The positivist paradigm is not appropriate for this research. Among others since the users opinions investigated are not an objectively given phenomenon.

The critical research paradigm views social reality as historically constituted produced and reproduced by people. Critical research could be used in this research if the focus were on identifying opposing conflicts and contradictions. And through the research help eliminate these. Critical research could also be appropriate if the focus were on the users' ability to change the overall system design. Focusing on the elements that are constrained by a domination of the projects interests tied to social, economic, legal and political issues. Some of these issues are touched upon in the ANT analysis of the research domain, but it is not the main focus of the research. The focus of this research is not to focus mainly on the overall

system design. The research is investigating the design of reference evaluation, which is just one part of the overall system design. The focus in the research of this thesis was on understanding the research domain through the eyes of the lab engineers and doctors working with the genetic classification. Thus, the interpretive paradigm is most appropriate, with a focus on understanding the specific context and how it influences the phenomenon being studied. In interpretive research, phenomena are studied by collecting the meaning that people assign them. This fits the research in the thesis, since the complex domain and work practice is studied by collecting the users' opinions. The system development of the article evaluation is studied through the eyes of the lab engineers and doctors working with genetic classification.

## 4.3 Mixed method research

The research findings in this thesis are based on a mixed methods research approach. Traditionally there have been a lot of discussions comparing qualitative research against quantitative research. Johnson and Onwuegbuzie (2004) present mixed methods research as the third research paradigm. Mixed method research advantage is in its methodological pluralism. The approach has the advantage of combining the strengths from quantitative and qualitative methods, effectively minimizing their weaknesses. The approach advocates that the researcher takes an eclectic approach at method selection. It also provides a more diverse viewpoint for reflecting about the research and how it is conducted (Johnson & Onwuegbuzie, 2004).

# 5 Methods

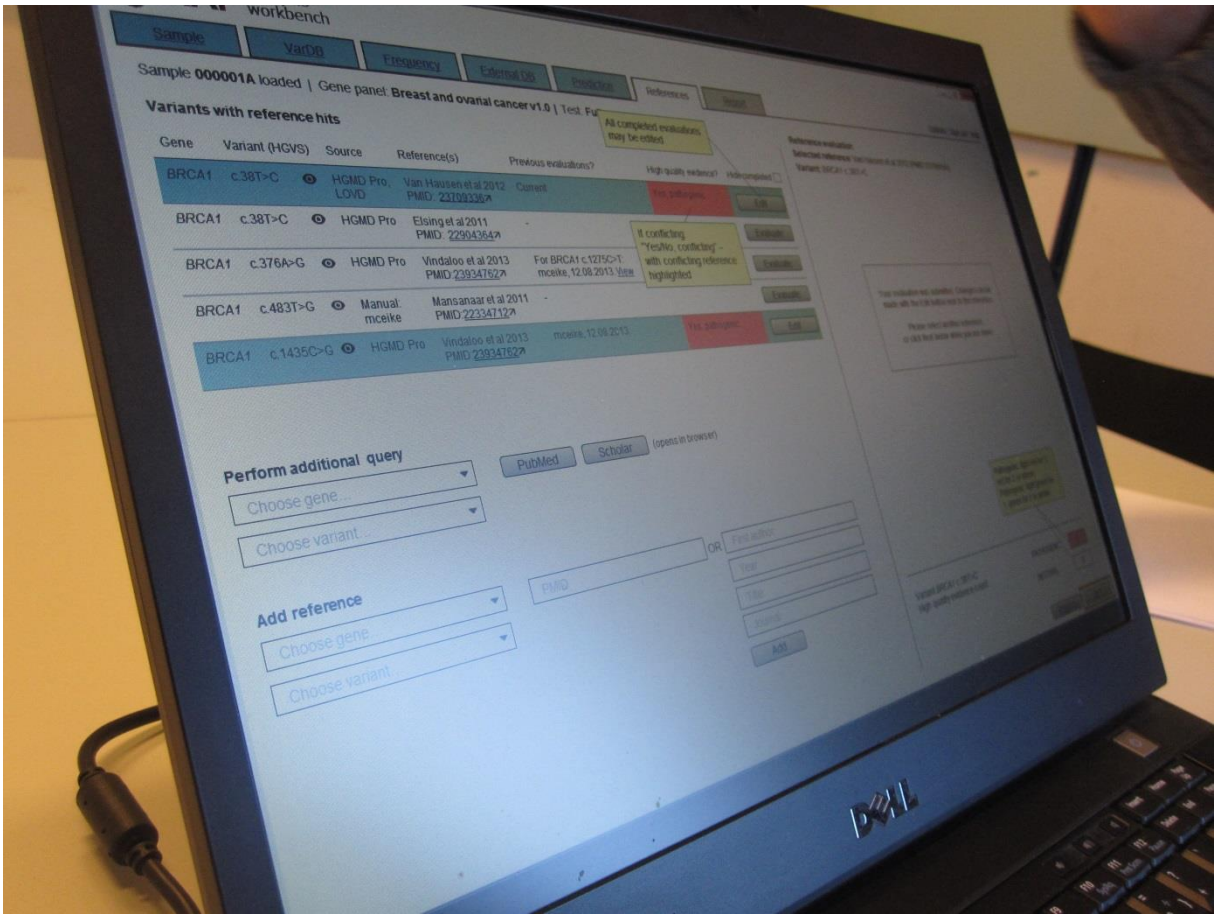



**Figure 3: Picture of the PowerPoint prototype.**

## 5.1 A prototype as a research tool

Research in HCI is often prototype-driven. Prototyping new system designs are used as research tools in the design process to explore new design concepts. A prototype with a developed specific design can work as the starting point for further design explorations (Wiberg & Stolterman, 2014, p. 538). According to Lim, Stolterman, and Tenenberg (2008) prototypes can have the purpose of manifesting design ideas and to provide a tool for the filtering of a design space. There are currently numerous HCI approaches for developing and applying new ideas in design. Like proof-of concept design, concept design, and different methods for working with design guidelines.

**Thesis research approach:** At the starting point of this thesis research a PowerPoint prototype of the system was already available, developed by the domain expert. This prototype was used in the data collection and was used as the starting point for further design explorations. The prototype was used in two usability tests of the prototype. Paper prototypes were also developed by adding functionalities with Photoshop to screenshot pictures of the

workbench being developed by the team of developers. These were used in the survey to collect research data.

## 5.2 Methods for gathering data

Data gathering needs to cover a wide range of issues to collect sufficient, relevant, and appropriate data. The methods are used to examine the tasks users currently perform and their associated goals, the context in which the tasks are performed, and the explanation for why things are the way they are. It is useful to deploy a combination of data gathering techniques to yield data triangulation, since each method results in data that is limited to a certain kind of information and perspective (Preece & Rogers, 2007, p. 219).

### 5.2.1 Observation

Simply asking users about their work practices is not sufficient for developing a complete understanding of the issues they encounter in their daily work. The users often have implicit practices that can only be collected through observing the users doing their work. Observations are especially useful to get an understanding of the context of a task performance (Preece & Rogers, 2007, p. 217).

**Thesis research approach**: In the fieldwork there were conducted two observations. These were conducted to explore how classifications of genetic variants are practiced by the laboratory engineers.

### 5.2.2 Interviews

Interview questions can spark reflections where the interviewees generates ideas and share valuable insights (Lazar, Feng, & Hochheiser, 2010, p. 178). Opportunistic interviews consist of taking an interesting idea and running with it. This can be useful for increasing understanding of topics or issues that emerge during the interview (Lazar et al., 2010, p. 179). This compelling flexibility could also lead to difficulties in managing potentially unbounded discussions. On the flip side if the discussion is to bounded by the researcher it can be difficult to capture the interviewees own unbiased perceptions and opinions of topics. To capture the users' opinions and their story described with their words, it is important to avoid leading questions (Madden, 2010). In the analysis and planning of the interview it is important for the researcher to be critical towards their role in the construction of the interview, and how that influenced the results. Since interviews are not natural occurring, but 'constructed' by the researcher it does not provide direct access to the experiences of the research subjects (David Silverman, 1998). Interviews give deep data that is not broad since it is often limited to one or a few participants. Conducting interviews require practice and experience to develop the ability of "moving conversation along, eliciting meaningful

responses, revisiting questions based on interview responses, interpreting subtle cues, and interpreting detailed responses"(Lazar et al., 2010, p. 178). To practice and identify questions that need further improvement it's useful to perform pilot tests before performing the interview. Interviews can be fully structured, semi-structured or unstructured. Fully structured interviews are a verbal presentation of a list of preplanned questions and there are no deviations from the list. These are easier to analyze and makes for a good comparison across individuals. Semi-structured interviews also have a list of preplanned questions, but digression is allowed. In this type of interviews the list serves as a starting point for a discussion. Unstructured interviews can start with initial questions and have a preplanned list of topics. There is no predefined script or topic for the interview so there is minimal structure. Generally the high structured interviews are easier to analyze and are good as a starting point for someone not experienced in performing interviews. The less structured interview yield advantages of open ended exploration (Lazar et al., 2010). In the beginning of an HCI research process interviews can be used for initial exploration of a topic to understand the needs and challenges presented in the context of a specific situation. Later in a HCI research process interviews can be used to gather data about specific details that would lead to concrete design requirements (Lazar et al., 2010, p. 181).

**Thesis research approach**: To explore the research question fully structured interviews are unsuitable for the data-gathering. The research questions require a deep exploration of the topics. Semi-structured interviews are best suited for exploring pre-planned topics and to investigate interesting topics that arise during the interview. Semi-structured interviews were used in the two usability test sessions of the prototype and as a stand-alone interview with a lab doctor. In addition semi-structured interviews were used after the two observations.

### 5.2.3 Document analysis

Newspaper article, research articles, TV documentaries, annual reports and other texts already in the public sphere can serve as empirical material (Finken, 2013a). Workplace rules and standards are often documented in manuals. These can be a good source of data giving background information about regulations governing work tasks and the steps involved in an activity (Preece & Rogers, 2007, p. 215). It is important to state that what users' do that is not in the manual might be very valuable for the researcher. For instance if the user deploy workarounds because of usability issues in a system they use, finding the reason for them might result in improved functionalities. "Taking a user-centered view of development means that we are interested in the everyday practices rather than an idealized account" (Preece & Rogers, 2007, p. 215). So document analysis should not be used as the only method.

**Thesis research approach:** During the first user session a document containing a literature evaluation and genetic variant classification was obtained from the user. This was a good source for analyzing a user developed strategy for research evaluation. The Standard Operating Procedure (SOP) documents containing the workflow and guidelines for article evaluation and variant classification was obtained from the domain expert. After the two

observations the variant evaluation schemas that were filled in were obtained. All these documents were a good source for document analysis.

### 5.2.4 Usability testing of prototypes

In broad terms usability testing involves representative users attempting representative tasks in representative environments on early prototypes of a system (Lazar et al., 2010, p. 252). The prototypes used can be paper based, look complete but have simulated functionalities («Wizard of Oz» technique), testing working versions of software before it is put into use or testing software that has already been implemented and are used in existing systems. The goal of usability testing is to improve the quality of an interface by finding flaws in it. The focus is both on identifying what that works well in an interface design and what that needs improvement. Such as major interface flaws that cause problems for many users. The usability test can be expert-based or user-based. The prototype needs to be user tested since the developers cannot clearly identify the users' needs, since they are not the user of the system.

**Thesis research approach:** As mentioned the prototype was used in two usability tests. The focus was both on improving the quality of an interface by finding flaws in it and to gather data addressing the research question. The sessions were also used to identify new or better design opportunities, together with domain expert who designed the prototype that was tested.

### 5.2.5 Think-aloud protocol

The think aloud protocol is a technique that requires the users to state out loud everything that they are thinking. By performing this technique their thought processes are externalized and the researcher can collect data about user's problem-solving strategies (Rogers et al., 2011, p. 256).

**Thesis research approach:** The think aloud protocol was both used in the two observations and the two prototype usability tests.

### 5.2.6 Survey

Surveys are a research method often used to collect data from many potential future users of a system. It can give a researcher an overview of a research population in a short amount of time. Surveys are usually non-intrusive and are therefore generally approved by institutional review boards (Lazar et al., 2010, p. 101). The data collected can be from a geographically dispersed population. It's important to choose the appropriate sampling method. The list of questions is formulated to elicit specific information from the responder (Preece & Rogers, 2007, p. 211). To develop a well formulated survey insuring the validity of collected data is time-consuming. It requires careful considerations to insure that the survey is well-designed and contains non-biased questions. To insure that the questions in the survey are perceived as

intended, it is important to perform a pilot test of the survey with a representative user. Since the surveys are often self-administrated (either on paper or by e-mail) they have to be designed to be easily understood and used by responders. Lazar et al. (2010) state that "most survey questions can be structured in one of three ways: as open-ended questions, closed-ended questions with ordered response categories, or closed-ended questions with unordered response categories"(Dillman, 2000; Lazar et al., 2010, p. 111). One limitation with the research method is that the data collected is not deep and detailed. Lazar et al. (2010) state that "if interesting phenomena start appearing, it is usually not possible to ask follow-up questions, or go back and change the original survey instrument to ask more detailed questions" (Lazar et al., 2010, p. 101). Surveys are often used together with other methods. The information and conclusions gathered in interviews can be further confirmed by a wider group through surveys. Users might have problems answering questions regarding usage patterns or moods due to recall bias. The data collected from surveys can be both quantitative and qualitative depending on the type of questions used.

**Thesis research approach:** A survey was planned and used to explore the topics and concepts identified in prior data analysis. The reason for using a survey was to gather data from multiple envisioned end users of the system simultaneously, and to validate the findings from previous user sessions.

# 5.3 Methods for data analysis

## 5.3.1 Content analysis

Researchers can due to new technology with low cost collect both audio and video recordings in addition to textual data (Lazar et al., 2010). These recordings from interviews, focus groups, usability sessions, diaries and observations can be used for content analysis. It is further argued that "text and multimedia information are great sources for researchers and practitioners to better understand their users or the interaction between their users and applications" (Lazar p. 286). Content analysis is a method used in many different domains like journalism, sociology, psychology and business. Although the method is frequently used there is not one unified definition of content analysis. Some definitions regard content analysis as applicable to both textual information and multimedia materials (e.g., drawings, music and videos). Other definitions restrict the method to be applicable only to textual information. The content analysis process uses both quantitative and qualitative techniques (Lazar et al., 2010, p. 285; Neuendorf, 2002). It is described as "a normally in-depth analysis that searches for theoretical interpretations that may generate new knowledge" (Lazar et al., 2010, p. 285).

Lazar argue that "text based data can provide information that can hardly be delivered through quantitative data". In an empirical study the researcher can collect quantitative data by measuring efficiency and accuracy. Typically this would be time spent to complete the task or

number of errors performed. One limitation is that this might not give the complete picture of issues and challenges experienced by the user. Since qualitative data does not cover the user's opinions of the application and suggestions for improvement (Lazar et al., 2010, p. 286). Such information can be very valuable in the design process and should be collected. One limitation with content analysis like with other qualitative methods, is that to generalize the findings is often difficult (Van der Velden & Culen, 2013, p. 191).

#### 5.3.1.1 Content analysis procedure

A priori coding and emergent coding is two different approaches to analyzing the data in content analysis. With a priori coding the coder predefines the coding categories. The categories are based on case relevant literature that describes established theories or frameworks. A priori coding is illustrated in figure 4.

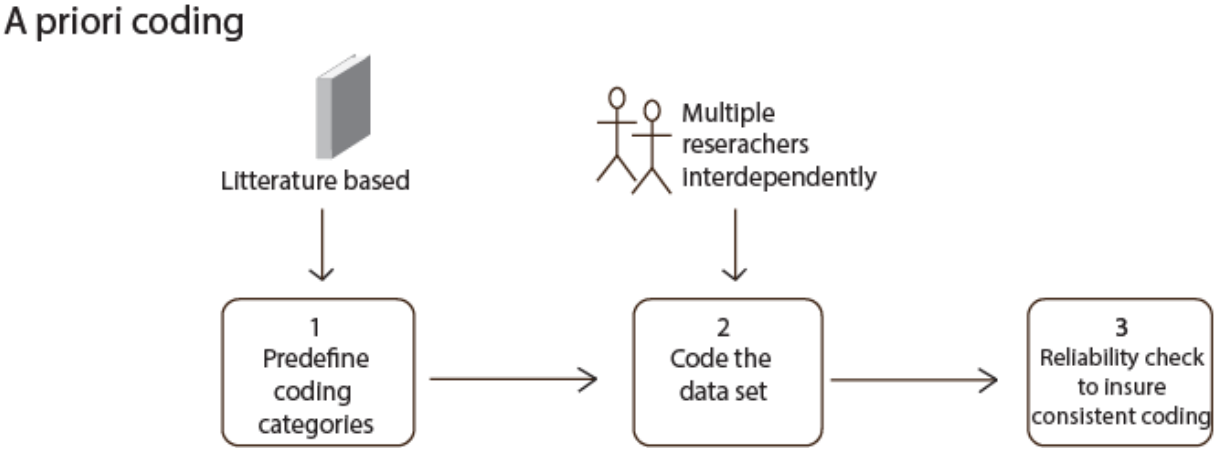



**Figure 4: A priori coding based on(Lazar et al., 2010, p. 289)**

Emergent coding on the other hand starts out with no predefined categories. The categories emerge from the data during analysis (Lazar et al., 2010, p. 289). Emergent coding is appropriate when there is limited literature or no establishes theories to develop the categories on. Emergent coding is illustrated in figure 5.

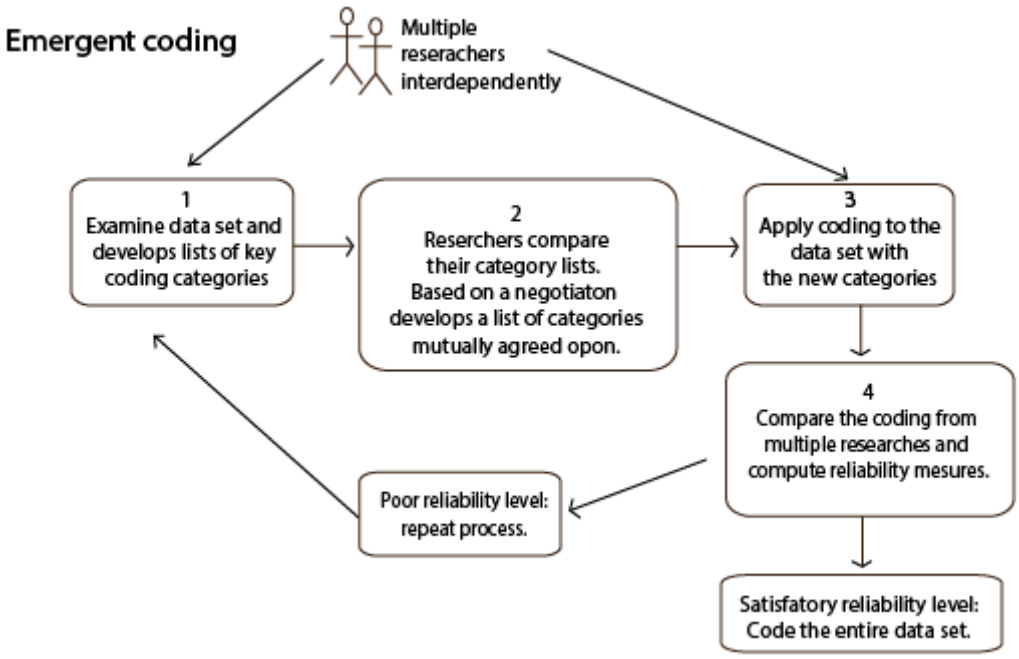



**Figure 5: Emergent coding based on (Lazar et al., 2010, p. 289)**

One problem in content analysis is that a coder might define categories that are highly subjective and that weaken the robustness of the results. Therefore two coders should code the data to ensure the quality of the categories. To ensure that the coding is consistent, reliability checks are performed after the coding task is completed (Lazar et al., 2010, p. 289).

**Thesis research approach:** The data from the survey was first sorted by placing all answers from one participant on one A4 sheet. The data set consisted of 11 pages and was coded by two coders interdependently using emergent coding. After negotiations between coder 1 and 2, the categories were agreed upon. A reliability check was performed. The emergent coding approach was most appropriate since there was limited literature and no relevant establishes theories.

## 5.4 Ethical considerations related to research

To ensure that the data gathering follows research ethics, it is important to use consent forms and insure anonymity and confidentiality of the participants. If research involves personal data, there are legal and ethical guidelines that need to be followed. There are legal requirements in Norway for how to process personal data in research. These laws further reinforce the ethical rationale for insuring the fundamental right to privacy. There are international regulations from UN Declaration of Human Rights, European Convention on Human Rights, EU privacy directive, The Nuremberg Code, The Belmont Report and The Declaration of Helsinki.

The Norwegian privacy law (POL) mandates that all data that may directly or indirectly be connected to a physical person must be reported to NSD. This concerns name, PIN, IP-

address, a localized patient profile of a rare disease. POL also regulates the use of sensitive personal data, which are data that reveals information about racial or ethnic origin, political opinions, philosophical or religious beliefs, health, sexual orientation and trade-union membership. The information that a person has been suspected of or charged with a criminal act is also regarded as sensitive data (Hannemyr, Autumn 2013).

**Thesis research approach:** Consent was collected from the participants before every user session, interview and observation. All the participants were informed that the data collection is confidential and anonymous. They were also informed that they have the right to withdraw at any time.

The survey (Attachment C) was emailed to the participants together with a consent form (Attachment B) and information about the research. Since the survey was sent out by email and answered by the responders through emails, their names would be visible for the researcher. Therefore due to the Personal Data Act § 31, it was mandatory to report the survey research to the Norwegian Social Science Data Services (NSD). The survey was reported to and approved by NSD.

# 6 Introduction to genAP Workbench and the initial inspiration phase

The initial phase of my thesis work was focused on reading literature and meetings with genAP project members, to get to know the genAP project and the domain of interest. Based on a discussion with the project manager of the genAP project it was decided that I could join in the user tests of a prototype for clinical genetic variant classification software named genAP workbench. In the design process of such a system it is crucial with some domain knowledge and a deep understanding of the design context. Throughout the research and data gathering my knowledge of the design context was constantly growing. To better understand and give an account of the design context of the genAP workbench and how it effect the development of the system the ANT framework was applied to the design context.

## 6.1 Design context

In the following section there is a presentation of the data collected about the case of the research. This presentation is shaped by applying the ANT framework to get a deeper understanding of the development process of the clinical genetic variant classification software. After this presentation of the context in whole, the focus is narrowed down to the smaller design context surrounding the reference evaluation functionalities.

### 6.1.1 Actor and actants in genetic classification processes

The process of genetic testing usually starts with referring a patient in need of genetic assessment to the clinic performing the genetic testing. There are two basic assessments that such clinics do and those are either diagnostic or prognostic testing. Diagnostic testing involves taking blood test samples from patients with a disease and testing if the disease is caused by genetic variant mutations in the patients DNA. Prognostic testing involves taking genetic tests from patients that do not have a disease diagnosis, but are at a high risk genetically, that is to say, patients that have close relatives who have, or had, a disease identified as having underlying genetic causes. Before taking the sample-test (a blood test) from the patient that is used for the genetic testing, the patients are called in for genetic counselling. If they proceed with testing, a sample is taken and, currently, processed through Sanger sequencing technology. The machine for sequencing has a certain capacity. For the genes BRCA1 and BRCA2 that are linked to breast and ovarian cancer the sequencing machine can process four different sample-tests together. The numbers of samples are based on what that fits into the concrete piece of technology, a machine used for sequencing. The LYNCH genes, which are linked to colon cancer (colorectal cancer), take less space, so the machine can test seven different samples together. After the test-samples are sequenced by the machine, the lab technicians receive a text-file with results called raw sequence data.

The raw sequence data consist of a unique combination of the letters “T”, “C”, “G” and “A”. Similar to the example code presented in figure 6, but much longer. These letters write the patients personal *DNA code*. The meaning of this code lays in the sequence of the letters like in regular languages.

```
TGTAAACCGGAGATGAAAACCTTTTTCCAAGGACAAATCAGAGAAAAAGTCTTTAACTCCACCATTAGCACCCA
AAGCTAAGATTCTAATTTAAACTATTCTCTGTTCTACATTTGGCCTCTACTTTTGGAAAAAGGTTCCTGTTTAGTC
TCTTTTTCAGAAATTGAGGTGGTAATCGTGGGTTTCGATTCTAAGATTAAATTTGATAAGAGACAAGATTCATG
GGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGC
CAGCCACCATGAATATTGTACAGTACCATAAGTACCCCTTCGTCTAAACCCATGGTGGGTTCATAACTGAGTGG
GTAGTTGTTGGCGATACATAAAGCATGTAATGACGGTCGGTGGTACTTATAACATGTCATGGTAAAATACTTG
ACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCTCCCCCCATGCTTACAAGCAAGTACAGCAATCA
ACCTTCAACTGTCACACATCAACTGCTTTATGAACTGGTGGACATCATGTATTTTTGGGTTAGGTGTAGTTTTGG
GAGGGGGGTACGAATGTTCGTTCATGTCGTTAGTTGGAAGTTGACGTGTGTAGTTGACGAACTCCAAAGCCAC
CCCTCACCCACTAGGATATCAACAAACCTACCCACCCTTAACAGTACATAGCACATAAAGCCATTTACCGTACAT
AGCACATTACAGTCAAATCTTCTCGTCCCCATGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCC
GTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAGAAGAGCAGGGGTACCTACTG
GGGGGAGTCTATCCCCAGGGAACTGGTGGTAGGAGGCACTTTAGTTATAGGGCGTGTTCTCACGATGAGAGG
AGCGAGGCCCGGGTATACACTTGGGGGTAGCTAAAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGCC
ATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGTGTGAACCCCCATCGATTTCACTTG
ACATAGGCTGTAGACCAAGGATGAAGTCCCGGTATTTCGGATTTATCGGGTGTGCAAGGGGAATTTATTCTGT
AGTGCTACCCGGGCATTCATCCTTGGTCATTGAGGTGGTAATCGTGGGGGCCCTGAAGTAGGAACCAGTAACT
CCACCATTAGCACCCTTATGAACTGGAAGCAGATTTGGGGCTCCGGATCCCGCACAAGAGTGCTACTCTCCTCG
```

**Figure 6: Example code of the structure of DNA code.**

The raw sequence data has to be quality checked and reduced down to a list of gene variants that are going to be analyzed. This process is challenging and requires high expert knowledge by the lab engineers’ performing these tasks. The data is cleaned up by removing artefacts that are known not to cause the disease. The material is then checked for some possible small deletions that might cause disease. If such are found, they are added back to the material for analysis. This is only a roughly explained description of the quality check; the actual process is far more complex.

Gene variant mutations happen when one of the DNA letters is accidentally swapped with another letter. A gene variant might have a very serious effect and be disease-causing, or have no effect at all. Generally, the variants are divided into pathogenic, neutral, or Variants of Unclear clinical Significance (VUS). There is no unified, agreed upon, classification system used to classify gene variants. The users who participated in this research use a classification system consisting of five categories, or classes. Class 5 variants are proven to be pathogenic, class 4 are likely pathogenic, class 3 are uncertain, class 2 are likely not pathogenic or of little clinical significance, and class 1 is not pathogenic or has no clinical significance (Plon et al., 2008, p. 1287) .

After the quality check of the sequence data, the lab technicians’ checks if identified gene mutations are normal. This is done by checking the variant list up against a paper sheet of variants prior classified as normal. Some of the normal gene variants are typical for Norwegians. These variants are called founder variants and they are removed from the list of

variants prior to analysis. After the list of variants is constructed, the next step is to use prediction software programs.

Currently, the lab technicians use Seqscape. Prediction software programs predict if a sequence is functional. The prediction information about the gene variants is used as a part of the annotating process. Currently the lab technicians use a program called Alamut, this program integrates many different software programs that are used in the annotation process. These software programs are LOVD, BIC, HGMD PRO, INSIGHT and dbSNP. The lab technicians also use in-house databases over prior classified variants. Then the lab technician use a search function in Alamut generating search strings in google covering all the different names used on a specific variant in the literature. Some articles are also found stored as prior evaluated references in the external programs LOVD, HGMD PRO and BIC. The research references are evaluated and if they are good enough they are used as a base for the resulting classification of the variant. During the annotating process the lab technicians fill out a variant evaluation form. The process is carried out by at least three people. First two lab technicians do the analysis individually. Then the lab doctor goes over the annotating evaluation and makes a classification assessment. The final report is then formulated and uploaded to a program called SWISS lab. The patient's doctor gets the results and forwards the genetic test result to the patient. If it is found a genetic cause for the patient disease this may affect the treatment.

The sequence technology used is Sanger sequencing, but the integration of HTS technology is coming, this will affect the current process of variant annotation. Currently the Sanger sequencing machine can handle 7 test-samples simultaneously for genetic testing of the Lynch genes. For BRCA genes the machine can simultaneously test 4 patients. For HTS this number Is 48 and this number are increasing as the sequence technology gets better. This will change the current practice of annotating gene variants.

## 6.1.2 An actor-network theory perspective on the development process

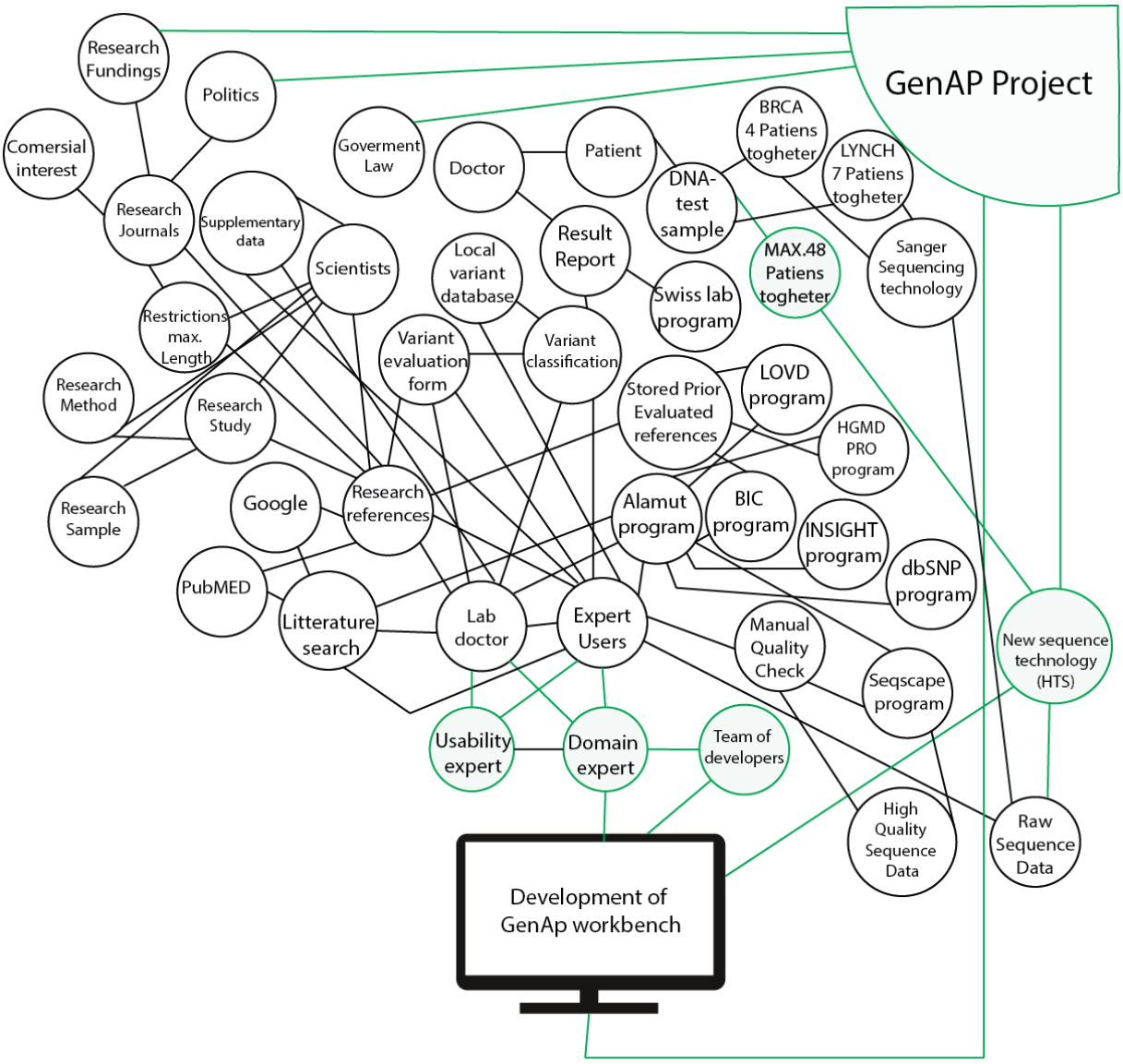



**Figure 7: Actor-network theory presentation of the case of research.**

Based on the case description in chapter 6.1, figure 7 is an actor-network theory presentation of the case of research. The green represents the relations between the actors and the intermediaries in the system development process. In ANT intermediaries has an important role, intermediary is anything that circulates between actors and helps define the relations between them (Gherardi & Nicolini, 2005, p. 4). In figure 7 the genAP Project has brought the system development (the green) into the already existing network (the black). The project is helping to define the relations between the team of developers, domain expert, usability expert and the users (user and lab doctor). The project has its own goals and seeks to impose its own version of reality into the network through the designers and developers, who are «acting on behalf» of the project. The project is achieving its goal (or effect) from a distance because it uses the intermediary that it has constituted. When the system is developed and integrated into

the network it could also function as an intermediary for the project. The project is also constituted by someone and is shaped by government law, politics and research funding's. The project might also function as an intermediary for these actors.

In the research in this thesis the usability expert (researcher) follows a user-centered design with participatory approach. Therefore it is possible to see the usability expert as an intermediary for the users (user and lab doctor). The development of the system is presented in 1.2.1, presented in 1.2 is the genAP project and the future introduction of HTS technology replacing Sanger technology that is currently used. The black is the current network before the introduction of HTS. The introduction of HTS technology will have a major impact on the current network, to accommodate the requirements and needs brought on by HTS is the projects main goal. Therefore it is possible to also see the project as an intermediary that is constituted by the HTS technology. If the designers use an expert mind-set in the design process the artefact that is the system, can also become the intermediary of the designers' desires, expectations and control efforts. However in ANT the system will get its role when it is integrated into the network and put into use and that is uncontrollable by the project and designers (Aanestad, 2003; Gherardi & Nicolini, 2005; Murdoch, 1998).

Just looking at the figure can seem overwhelming and it is evident how complex the development process truly is. The users have a crucial placement and hold the knowledge of how the work of genetic classification is done. This is one important argument for involving the users in the design process. For the experts to maneuver in such a complex network successfully, it requires both expert domain knowledge and high amounts of tacit knowledge. To gain an understanding of the tacit knowledge used, it is valuable to apply HCI research methods. The further need for domain expertise can be strengthen by the usability expert working together with a domain expert with usability interests in the user sessions (Chilana et al., 2010, p. 2337). Teamwork between a domain expert and a usability expert was used in most of the user sessions in this thesis's research.

From an ANT view all the relations presented in the figure 7 is important for the development of the system. It is important to state that the actors who are included in this network are limited to the issues and elements addressed or touched upon during the research in this thesis. The related project and development as a whole is much bigger. How the infrastructure challenges are handled in the project is outside the scope of this thesis. What is apparent from this figure is that the shift from Sanger to HTS will give big challenges both infrastructural, on the users, and on programs. The representation is also restricted to the time of research. The network is evolving and changing constantly through the development process.

#### 6.1.2.1 The heterogeneous network

A crucial concept in ANT is heterogeneity. All the networks in ANT are viewed as heterogeneous of "diverse (not simply human) materials" (Law, 1992, p. 380). The actors in the development process of the software are characterized by their differences that affect the network. The magnitude of these effects is dependent on the differences. Therefor ANT can

be defined as a framework used to understand a heterogeneous network consisting of diverse human and non-human actors. In the development process there is a need for humans (like the developers, domain expert, users and patients) and various non-human actors (like paper schemas, external software programs and research articles). All the different actors have a role and some sort of influence in the network representing the development process. The differences in patients' genes will give different results of the genetic testing. Currently some test-samples will not give a list of genes for analysis so there will not be a gene annotation process. Some test-samples will result in many gene variants. Some gene variants will have prior classifications and some gene variants will not be well-documented in research literature making them difficult to classify. There are several different other scenarios which will affect the role of the different other actors in the network. The level of difficulty is highly fluctuation dependent on the different cases brought on by the differences in patience genes. This illustrates how the genetic testing process is a heterogeneous network. The actors will have different roles in the network depending on the differences in the patients' genes. The patients are also a heterogeneous actor. The development process has to consider all these different cases. The knowledge of how these cases are currently handled is crucial and is gained through the relation with the users.

The lab doctors and the users bring their different competencies and experiences into the network. Similar the domain expert, developers and usability expert also bring their different competencies and experiences into the development process. And through the user involvement in the development process all these different competencies and experiences effects the development process.

#### 6.1.2.2 A priori

Aanestad states that "the theory argues that it is analytically fruitful to reject any *a priori* distinction between elements in the network, as e.g. the distinction between humans and non-humans" (Aanestad, 2003, p. 7). When doing research of the genetic testing process and the actors involved in the process it is important to focus on how the roles of the actors is defined in the specific network. One example is how Google is used to find research references and is an actor in this network. Google is used with many other purposes and the researcher might have previous established personal opinions about this browser. These personal opinions should not affect the definition of the actor's role in this network. The role of Google is constructed in the network and the interaction with the other actors. That role is what the researcher should focus on.

#### 6.1.2.3 Agency

In ANT the actors has an agency. Aanestad states that "agency is an *emergent* and not an essential or inherent property of the actors. It is the network that provides the actors with opportunities to establish and use agency" (Aanestad, 2003, p. 7). An example of this is the

mentioned role of the Google browser in different user cases. For instance Googles role will emerge different if it is used to search up the phone number and contact a long lost friend. The research articles might have been written by the scientist just for scientific reasons. But in this network the research articles are used as a base for diagnostic decisions that have impact on patients' life and health. This changes the agency of the actor.

### 6.1.2.4 Delegation

In a network the actors can be delegated other roles. Delegation occurs when an actor take over another actors role and act on behalf of this actor (Aanestad, 2003, p. 7). One example of this is if the external software programs are replaced with other programs. Or if the manual quality checks off the raw sequence is computerized. Through the development of the clinical genetic variant classification software the role of the Alamut program is changed. Currently Alamut has the role of integrating some external software tools and is much used in the annotation process. The clinical genetic variant classification software is going to integrate all external software programs and the role of the Alamut program is going to change. Through the delegation the new software also gains the power, competence and responsibilities currently associated to the role of the Alamut program (Aanestad, 2003, p. 7).

### 6.1.2.5 The alignment of the network

In the ANT theory symmetry is important and is accomplished by equating the human and non-human actors. The symmetry is achieved when all actors are aligned. This means that the actors more or less share the same goals and motives. Aanestad (2003) states that "The alignment of the network occurs through a process where the actors' interests are translated (i.e. reformulated, modified, or changed) into more generally agreeable expressions, so that several actors may support the resulting translation" (Aanestad, 2003, p. 7). This translation may involve inscriptions like rules, standards or the characteristics of an artefact (Latour, 1990). The purpose of the translation is to define a program of action. The program of action is a proposed framework of varying strength that defines which actions that are possible. Aanestad (2003) states "if the actors are supporting (more or less voluntarily) a given translation and its inscription, they have been enrolled and are cooperating (still more or less voluntarily) towards a common goal" (Aanestad, 2003, p. 8; Latour, 1987).

The Clinical genetic variant classification software's structure is based on the users workflow this can been seen as an inscription. There might also be a need for using constrains in the interface that is used in the translation as inscriptions. In the network the actors adds their knowledge and abilities, to get stability the various actors have to comply with the other actors. The knowledge, ability and opinions from the users have to be identified in the development process, to be able to translate the experts' interests. Further there have to be a process of compromising if the interests from the stakeholders and the users are different. To make the network stable and aligned the result from the translation need to be agreed upon by

the actors. This is another argument for the need for involving the users in the development process.

The finalized design of the Clinical genetic variant classification software and how the network is going to get its role in the network is not pre-given. The software gets its role through use. How the software is used cannot be defined solely by the designers. If the system is used in another way than the planned goals and objectives it is called drifting. The notion of drifting also covers the possible unintended consequences of introducing information technologies (Aanestad, 2003, p. 1). User strategies are deployed and the software might not be used like intended. For an instance if the users deploy workarounds to overcome solutions in the software that does not suit the user. In the comment field used to evaluate research references if the text is easily lost without being retrievable. The users might start using other text processing programs with a save functionality and then pasting the completed comment into the comment field. Then the software would not have the role intended by the designers. One of the roles intended for the comment field will be delegated to an external actor that is integrated into the network by the users. The users also deploy different strategies to maneuver in the complex network successfully. The development process has to integrate support of these strategies to "create an interrelationship between the technology and the work practice" (Aanestad, 2003, p. 2). The human activity of the users has to be researched and translated into design solutions that both stakeholders and users agree upon. This is achieved through negotiations between the actors in the design process and by focusing on how the "design will happen more as design-in-use and less in defined design projects" (Aanestad, 2003, p. 2).

One of the big challenges in the development process is to foresee how the integration of HTS sequencing technology will affect the network. Many of the actors in the network will be affected. Both the human and non-human actors will be affected. The current actors' roles will either be removed or reestablished in the network. Some of the used external tools will probably be exchanged. New actors will be included in the network and there will be a big change in the workflow of the users and a massive increase in workload.

As mentioned in the start of this section the issues of the infrastructure surrounding the development of the genAP system is outside the scope of this thesis. These issues are addressed by others. The focus is on how the reference evaluation functionality that is one part of the system should be designed based on user needs and preferences. There will be many more references to evaluate as the amount of gene variants per analysis and the amount of scientific research increases. Still the experts' users strategies for evaluating research references will probably be one of the few roles in the network that is kept somewhat the same. It is however important to consider that from an ANT-view nothing happens in isolation. The reference evaluation will be affected by the other actors in the network. Based on the figure it is apparent that the research references are affected by the scientist(s) performing the scientific research and author(s) of the article. The research journals restrictions of maximum length further influence the article's content. Other actors influencing the scientist is commercial interests, politics and research funding. Users evaluating the research articles are focused on extracting information about the study in the article covering

the specific gene variant. The information needed for using the reference as a base for a medical diagnosis, is especially the study's material and method. The interaction between the scientist and the user is mediated through the article, which is an object made of paper. This is a non-human actor that has an important role in the network and participates in the shaping of the interaction (Law, 1992, p. 382). The article mediates the communication between the user and the scientist. The mediation is done asymmetrically, amplifying what the scientist is communicating through the text without giving the user an opportunity to ask questions or participate with their opinions (Law, 1992, p. 382). Therefor for the article to be valuable for the user, the descriptions of the studies have to be well-documented in the article text. The study can be performed well and give high indications for the validity of the study results. However, if the study's material and method is not well enough described in the text, it cannot be used by the users for the classification of a variant. One of the users stated in an interview that "I'm interested in more detailed information about the gene variant. Information that I think that they most certainly have, but that is not presented in the article. But this is caused by the conditions restricting the number of maximum words that are allowed in each journal. Sometimes one's lucky and finds supplementary online tables."(from user session) The use of supplementary data is one example of an element that is easily over looked by a priory assumptions. This a-priory assumption would be that the evaluation of references solely consists of evaluating PDF articles. If supplementary data is not integrated as an actor in the network the development of the system might only support the uploading of files with PDF formats. And if the online supplementary data suddenly is unavailable, the user might not be able to use the reference in the classification process.

To sum up, the user strategies deployed to evaluate research references is important to identify. This is an argument for using HCI- research methods and involving the user in the development process of the software. These strategies need to be translated and addressed in the design process to mediate the design-in-use process that starts after the software is put into use.

## 6.2 Context "in small"

In 6.1 the context surrounding the workbench system was explored. The user studies in this thesis focus on a smaller part of the context. More specifically, the future users of genAP workbench and the prototype of the system.

### 6.2.1 The prototype for references

The genAP workbench is presented in section 1.2.1. The workbench and the users of the system are also affected by the larger context. The aim is to get a more real understanding of the users' situation in such a complex domain. It is important to involve the users in the design process and perform observations at their place of work. The observations performed in this thesis identified opportunities for design. That was not part of the prototype. The

observation that the users search the PDF's for the variant names was uncovered. By semi-structured interviews and a survey, it was brought forth that many do search PDF's for key words. This shows the importance of also looking past the prototype system focusing on the current context of the future users work practice.

### 6.2.2 The future users of the system

The laboratory engineers working with the genetic testing are users with high levels of expertise; they possess both complex domain knowledge and tactic knowledge. In this research the users are seen as the expert on the design context and how to design for reference evaluation. Their knowledge informs the design through the different user studies.

# 7 Data gathering

In this chapter the data collection is described, starting with the planning of the data collection.

## 7.1 Planning the data gathering

The data gathering was planned to insure that the data was obtained correctly, was relevant and sufficient for addressing the research questions. The various sessions with the users were planned in prior meetings.

## 7.2 Participants

The participants attending the user sessions are all future users of the system. They are from a small user group consisting of users with expert knowledge combined with domain knowledge. Therefore they are all highly representative users for the system. One of the participants was a lab doctor the others were laboratory engineer. There are some differences between the users in specializations within the fields of expertise. For instance some work more with one type of cancer testing like the BRCA-genes, while others work with different genes. Through the data collection users with different specializations in the field of expertise were included. There were more female participants, but since there are a majority of females in the user occupations. This was still seen as representative for the user group.

## 7.3 Post session meetings

After all the user sessions the data were transcribed and analyzed as part of this thesis. After the data were analyzed there were meetings with the domain expert. The results were presented and discussed with the domain expert. Some of the results were addressed by changing the design of the prototype, others were further investigated in later sessions.

## 7.4 Data gathering sessions

The data gathering consisted of different ethnographic methods, a survey and two think aloud usability sessions with the prototype. Most of the sessions were audio recorded, some were video recorded. The total data gathering process is presented in Table 2.

| Name | Focus | Method | Data | Duration | Participant() |
|---|---|---|---|---|---|
| User session 1 | To identify the most challenging tasks and gather data relevant for the research question. Identify usability issues. | Talk aloud, Usability testing of prototype. Semi-structured interview | Audio recording, some video, pictures. Note taking. Obtain document containing literature evaluation. | 1 hour 40min | 1 laboratory engineer |
| User session 2 | To identify the most challenging tasks and gather data relevant for the research question. To further explore and validate issues identified in user session 1. Identify usability issues. | Talk aloud, Usability testing of prototype. Semi-structured interview | Audio and video recording. Note and picture taking. | 60 min | 1 laboratory engineer |
| Observation 1 | To gather data about the reference evaluation functionalities. | Observation followed by , semi-structured interview | Audio recording. Note and picture taking. Obtain document filled out during the task of variant classification. | 2 hours | 2 laboratory engineers |
| Observation 2 | To gather data about the reference evaluation functionalities. | Observation followed by , semi-structured interview | Audio recording. Note and picture taking. Obtain document filled out during the task of variant classification. | 1 hour 12min | 1 laboratory engineer |
| Interview | To gather data about the reference evaluation functionalities. | Semi-structured interview | Audio recording. Note taking. | 1 hour 45min. | Lab doctor |
| Survey | To gather data about the reference evaluation functionalities. | Survey sent out by mail to future users of the system at the hospital | The word documents containing the survey answers. | | 11 participants |

**Table 2: Presentation of the total data gathering process.**

### 7.4.1 User session1

The first user session was conducted to explore the prototype, and address the research questions mainly “which tasks are most challenging when using the clinical genetic variant classification software?” Data relevant for the design implications of the task identified as most challenging was also collected. The session was conducted at the research hospital with a domain expert and two usability experts. The participant was a laboratory engineer working with variant classification related to cancer genetics and is a future user of genAP workbench. The session was planned in a prior meeting between the usability experts and the domain expert. The session started with some questions collecting some initial user expectations and

needs. The rest of the session consisted of a usability test with the prototype using a think aloud protocol and performing a walkthrough of the prototype. There were some questions used prior to introducing the prototype to collect initial user expectations and needs. The prototype walk through was followed by a semi structured interview. The session lasted 1 hour and 40min, the audio recordings were transcribed into 11 039 words. After the session was completed concerns were raised that the results from this session was not representative for the general user group. Therefore a second user session was planned and executed to insure the validity of the results.

#### 7.4.1.1 Document analysis

From the first user session there were obtained a document from the participant containing a variant classification with literature evaluations. This document was analyzed and the results were later compared to the other results.

### 7.4.2 User session 2

The second user session was conducted similar to user session 1. The focus was still to explore the prototype and address the research question. This session was conducted at the research hospital with a domain expert and a usability expert. The participant was a laboratory engineer working with variant classification related to hereditary heart disease and is a future user of genAP workbench. Similar to session 1, it was planned in a prior meeting between the usability expert and the domain expert. The session started with some questions collecting some initial user expectations and needs. The rest of the session consisted of a talk aloud walk through of the prototype followed by a semi structured interview. The issues encountered in user session 1 were addressed in the questions to investigate how these issues were perceived by a second participant. The session lasted 60 min, the audio recordings were transcribed into 7 701 words.

After the analysis of the collected data from user session 1 and 2, the initial research question were answered. The focus was now on research references and the research question was "How can the use of user-centered design and user participation, considering EKP (emerging knowledge processes), design theory and actor-network theory framework as important in the complex context of genetic analysis platform design, support the design process of clinical genetic variant classification software?" In the following observations and interview another researcher was involved. The researcher was investigating how the current work of variant classification is performed in the hospital. The data collected from these sessions used in this thesis was primary related to evaluation of research references, but some data were also used to increase the knowledge of the domain and design context, presented in chapter 6.

### 7.4.3 Observation 1

To further investigate how articles are evaluated there were set up an observation. The observation was done in collaboration between two researchers. Audio recordings, pictures and note taking were performed during the session. The participant was a laboratory engineer working with variant classification related to cancer genetics and is a future user of genAP workbench. Similar to session 1, it was planned in a prior meeting between the researchers. After the session was completed notes and pictures were exchanged between the researchers. The audio recording was 120 minutes long. A total of 32 pictures were taken.

#### 7.4.1.1 Document analysis

From the first observation there were obtained variant classification documents, containing the observed classification process. These documents were analyzed and the results were compared to the other data from the observation.

### 7.4.4 Observation 2

To cover more user groups a new observation was planned. This observation was conducted by a researcher, a domain expert and a usability expert. The participant was a laboratory engineer working with variant classification related to genetic heart disorders and is a future user of genAP workbench. Similar to session 1, it was planned in a prior meeting between the researchers. Audio recordings, pictures and note taking were performed during the session. After the session was completed notes and pictures were exchanged between the researchers. The audio recordings were 72 minutes long. A total of 7 pictures were taken.

### 7.4.5 Interview with laboratory doctor

This interview was conducted by a researcher, a domain expert and a usability expert. The session was conducted to explore how the lab doctor is involved in the variant classification and article evaluation.

### 7.4.6 Developed paper prototypes to be used in the survey

Based on the analysis of all the data collected, questions were made and two paper prototypes were developed to investigate the research questions further in a survey. The prototypes are Photoshop redesign of a high fidelity prototype that was being developed. The new system was being developed by the team of developers, using the Python programming language. It was based on the initial PowerPoint prototype and the results from the user sessions.

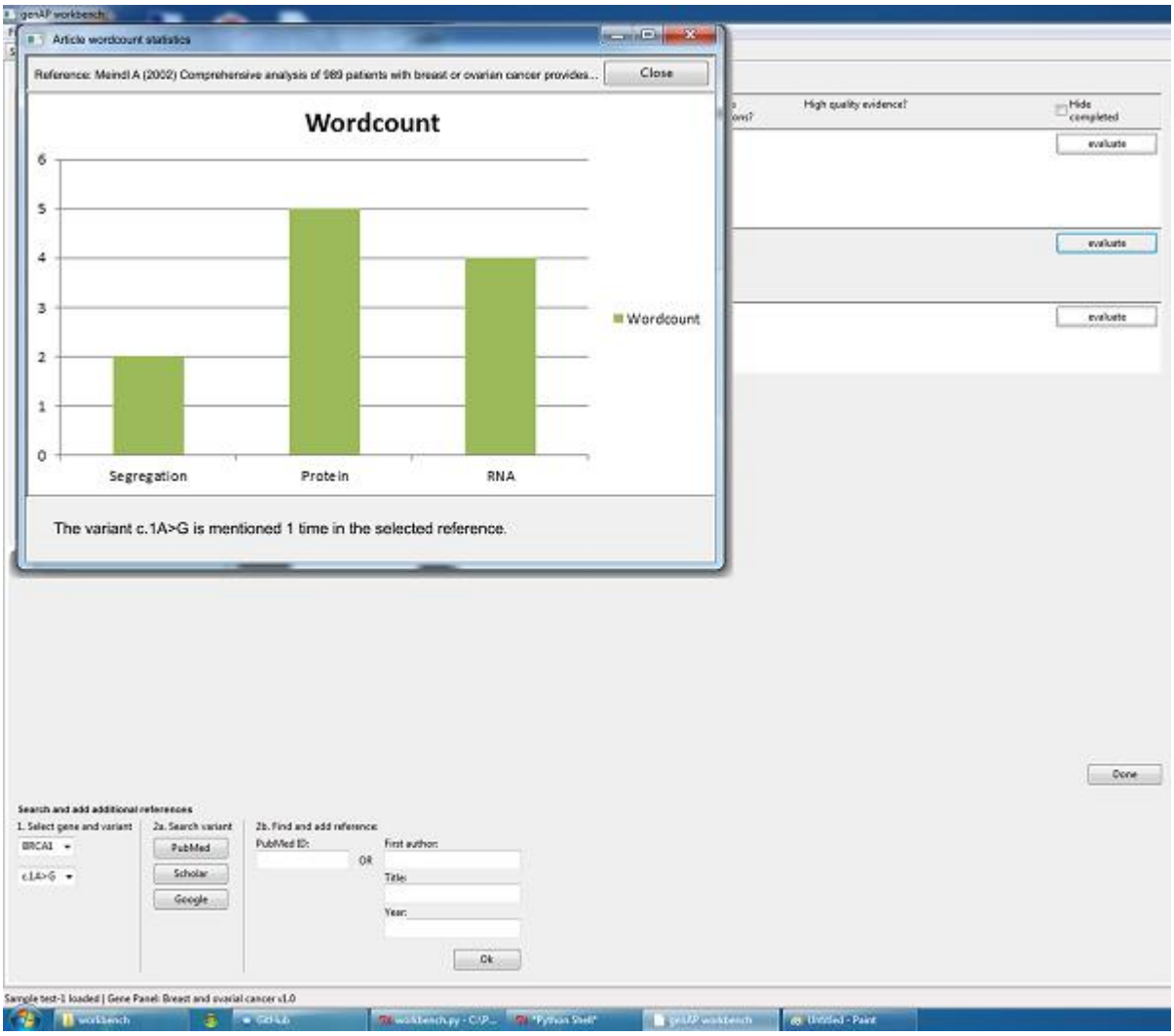



**Figure 8: Prototype displaying the possible word count functionality.**

The functionality added in Photoshop is the box displaying the graph showing “word counts” of identified key words in the article. This was developed to support the users in their decision process to decide which article to evaluate from a potentially long list of articles. The list is behind the box, from the list the user can select an article and then select to see the box displaying word counts related to the sleeted article.

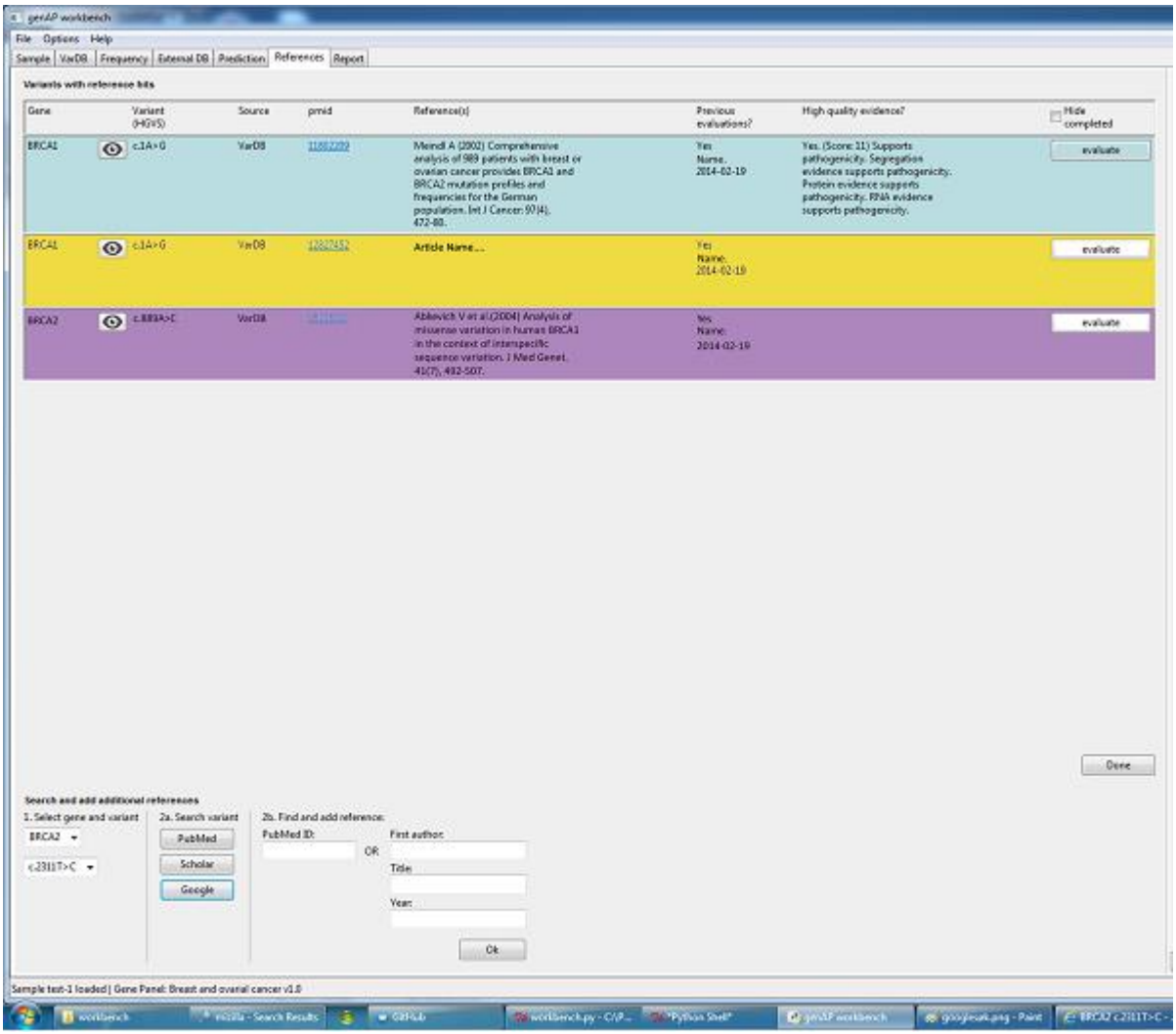



**Figure 9: Prototype displaying possible color coding.**

In this prototype the yellow and purple color is added to highlight selected articles by prior evaluators. Yellow are articles marked as containing issues so it signals to use caution with these articles. The purple are articles selected by prior users as favorites that are especially useful in the classification of the specific variant being classified.

### 7.4.5 Survey

To further investigate the research question a survey was made. The survey was developed and then it was discussed with the domain expert. After further improvement it was e-mailed back and forth between the domain expert and usability expert for a period of 4 weeks, a process consisting of discussions, changes and improvements. During these weeks the survey was also reported to the Norwegian Social Science Data Services (NSD) and was awaiting approval. After the survey was completed a pilot test was performed by sending the survey to a participant that performed the survey. After the survey was completed by the participant a meeting was used to discuss the survey and improvements were done before sending the

survey out to a total of 18 participants that were all future users of the system. After 16 days and a reminder the survey was answered by 11 participants.

# 8 Analysis and data

**Figure 10: Picture from the analysis of the gathered research data.**

## 8.1 Results from the first phase of research

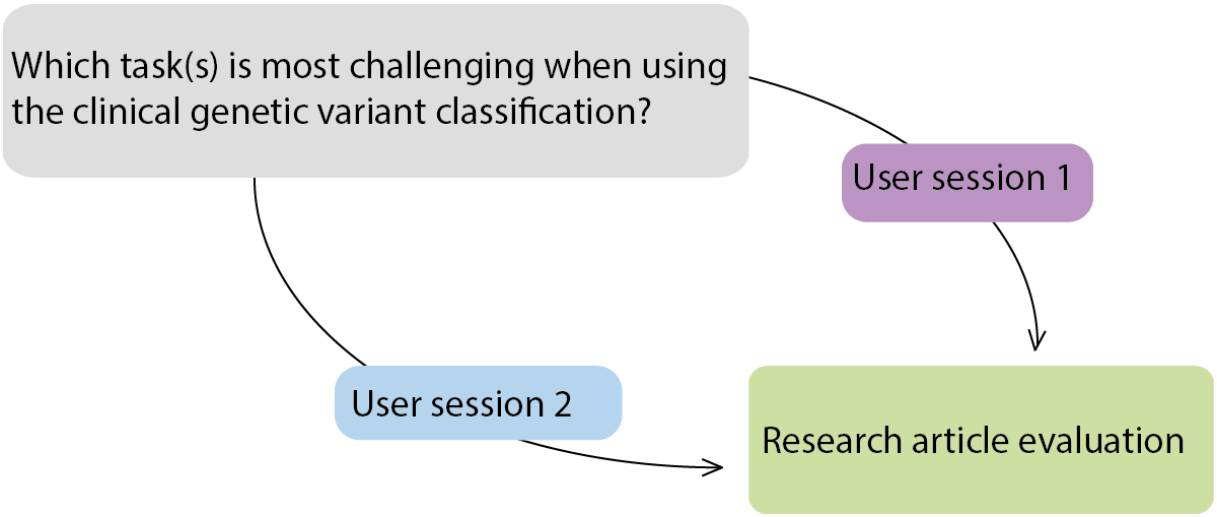



**Figure 11: Illustrating the first phase of research.**

### 8.1.1 Results from the first two user sessions

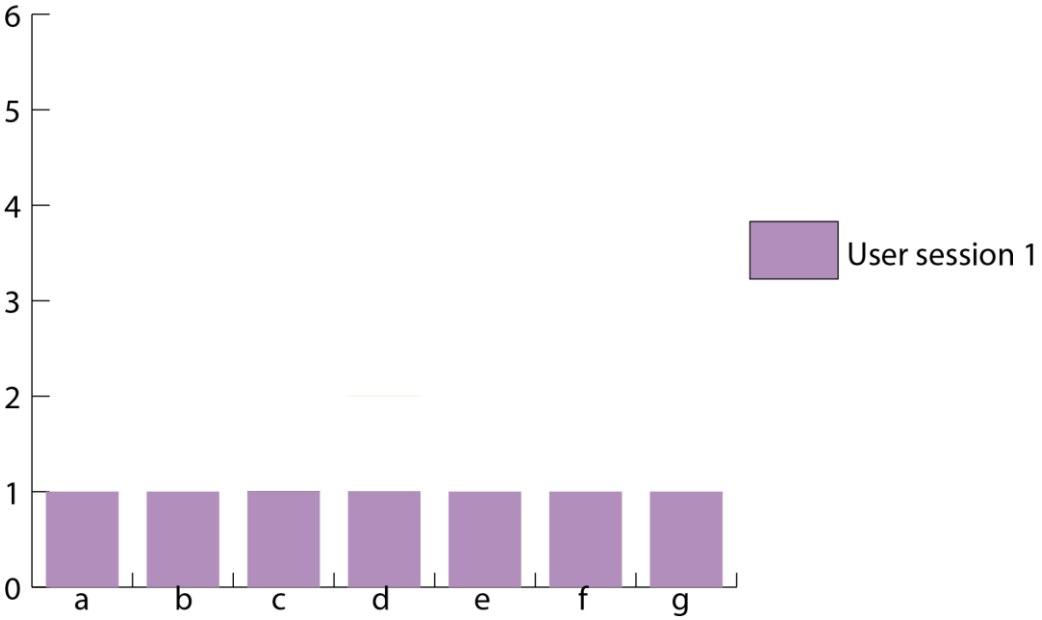
The first two user sessions were deployed to investigate the two sub research questions regarding the case of research. These are "which task is most challenging when using the clinical genetic variant classification software?" And "what are the important design implications for the functionality of the identified task?" Through the analysis, there were identified two tasks that were challenging when using the clinical genetic variant classification software. The first were the task of constructing the list of gene variants to be analyzed in the program and getting the data into the program. This consists of quality checking, removing technical artifacts, identifying small deletions and removing the variants that are normal. This task is currently outside the scope of the prototype. The second task was perceived as even more challenging. In the user sessions both the users state that the evaluation of references is the most challenging part of the annotation process. Therefore, it was decided that I should focus on the task of evaluating research references.

In these two user sessions there were identified usability issues and user preferences regarding the functionalities in the prototype. Some of these were used by the team of developers working on implementing genAP workbench. For a more detailed description of the results from these user sessions please look in Attachment A. Through the user sessions with the prototype there were gathered other information that were analyzed and used in later stages of the research as well.

## 8.2 Findings related to the design of the research article evaluation functionality

When it was decided to shift the focus to reference evaluation, there was a need to go back and reanalyze all data focusing on this specific topic. There were several problems regarding the design of the reference evaluation identified in the two initial user feedback sessions. The challenge after the new round of data analysis became that careful inspection gave even more issues to consider. Some issues where briefly mentioned, while other issues where recurring and repeating as a "red thread" through the data. Issues that were briefly mentioned were also important, but considered by users as easy to resolve. The graphs in Figures 12, 14, 15, 17, 18, and 25 are made to show themes that were important in all observations and user sessions and make this "red thread" visible also to readers of the thesis. The survey then was designed to further study seven issues, this time including all engineers and a lab doctor, providing an opportunity for all to influence the design.

## 8.3 Findings from the first user session




a. Multiple user strategies for deciding which article to evaluate first.
b. Describe study type in list of articles.
c. Color coding of evaluated references.
d. Possible to mark/register issues of uncertainties.
e. Paste quotes from article into comment field together with own comments.
f. Difficult to extract the information in article about the specific variant.
g. Describe study type, material, method in comment field.

**Figure 12: Graph displaying the identified "red thread" issues and themes, from user session 1.**

### 8.3.1 Themes and issues identified

There are multiple user strategies for deciding which article to evaluate first. When conducting an evaluation of an article, it is difficult to extract the information in the article that is about the specific variant. Since many variants are often examined in an article with different study types. To identify the study type used on the variant is important for the variant classification. The type of study, study material and study method used on the specific variant is described in the comment field. The user states that she likes color coding of evaluated references and would like to see more color usage in the overall system design. She states that they prefer to paste quotes from the article into the comment field together with own comments. The user would like to mark or register issues of uncertainties, if she is a bit uncertain of something and wants the next person to contribute with their opinion about the issue.

### 8.3.2 Differences in assessments

When asked what is the most challenging with the task of reference evaluation. The user states that the problem is that people asses the same reference differently. Therefore it is important that one can see why someone has made the choices they have. She further states that «It is a demand that the system supports the communication of the previous evaluators' thoughts and reasons for their assessments». She exemplifies that if the decision is made because the article contains segregation data, then this should be conveyed to the next person doing the evaluation of the reference. It is often hard to asses based on the articles content if the article contains segregation data. She further states that "the system must insure that one understand what others are thinking when they make their assessments."

### 8.3.3 References in the final report

The references are put in the report, perhaps only a few selected ones. She wants to be able to choose how many references that are put in the final report. And make sure that the ones that are included are only the ones that were important for the evaluation. Type of study in the reference used on the variant should also be displayed in the final report.

### 8.3.4 Quality check

When articles are found they are uploaded into the workbench for use in the analysis. Related to this the user states that "it must be a check of how you type in so that it is stored correctly. It can be many authors and the systems most register them correctly". There is an issue in currently used systems that people mistype.

### 8.3.5 The lists of references

The user states that “to display all the references without some sort of ranking is not necessarily the best option, as there might be as many as 20 articles in the list.” In the prototype the user can select the variant and search and then it is generates a search string that can be used in a Google Scholar browser. The user likes this functionality and states that this is similar to a functionality she values highly in the Alamut program. The users request an added button for regular Google as well. The prototype is designed to have the PDF’s of the reference articles available in the program and accessed directly. The user finds this valuable since there might be problems getting access to other areas where the reference is initially accessed from. This is relevant for the ease of use of the program.

### 8.3.6 Reference evaluation schema

In the prototype it is used a schema where the user selects the correct options for the reference article that is evaluated and the workbench calculates a score and recommendation based on the score. The schema is developed to be used as a guidance tool. The user states that the schema and comment field in the prototype might work well, but that she does not know without having used it. She states that “I feel unsure to select this is segregation data if I'm not absolutely certain.” She further predicts that most likely uncertain users will handle it by not ticking the box and states that that in a sense that would be wrong. When asked if the type of study is covered by the schema, the user states “I don’t know.” This issue is important to address in later user sessions.

### 8.3.7 Emergent demands

As the research of gene variants progresses there are probably going to be cases where references should be linked to many gene variants. The system has to handle these special cases. The user illustrates this by stating” we had a disease-causing gene variant in gene codon 335 a stop codon and it was classified as disease-causing since it was well documented. And then we had another variant in stop codon 336, the next codon and it was not well documented, so it was only classified as a class 4 gene variant. But you can imagine that if you have one in 335 that is disease-causing, why should not it be that in 336?” She further states that If we just get that one good article that states that «here this gene domains cannot withstand anything.” She wants it registered in the system and linked to all related gene variants.

## 8.3.8 User developed strategy for evaluating research references

The user has brought two paper sheets containing a user developed strategy for evaluating research references for gene variant classification. The user in the first user session has brought a paper sheet of a method developed by the users to evaluate references. She states that this method was perceived as a good way to work together and handle these uncertainties through collaboration. The first person typed in their assessments of the articles and noted in the text if there were any issues of uncertainties. Then the second person read and built the text further. The aim of using the comment fields is to emphasize to the next person what ones thoughts were. Following is a picture of the two pages for reference evaluation after the completion of the assessment of a gene variant classified as a class 5 based on the evaluation of two references.

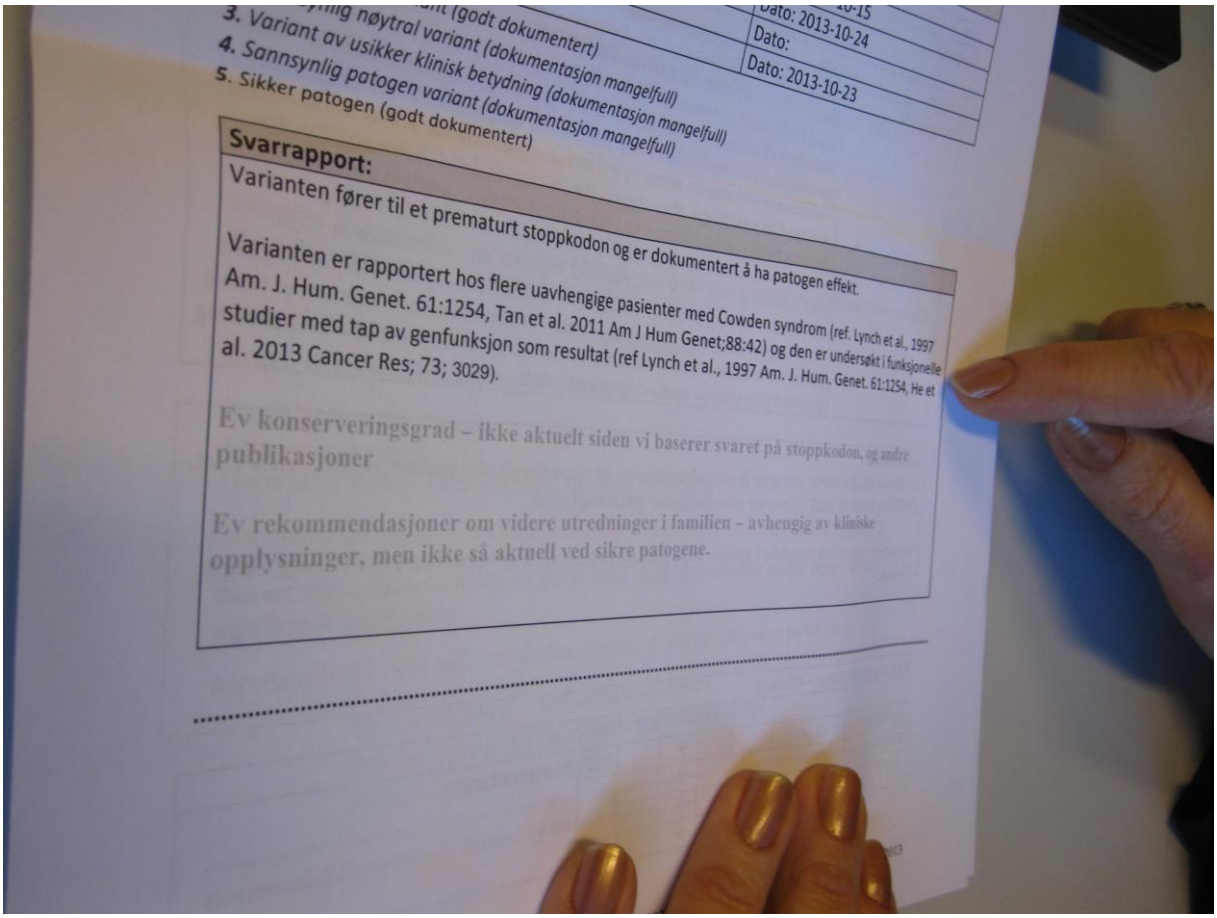
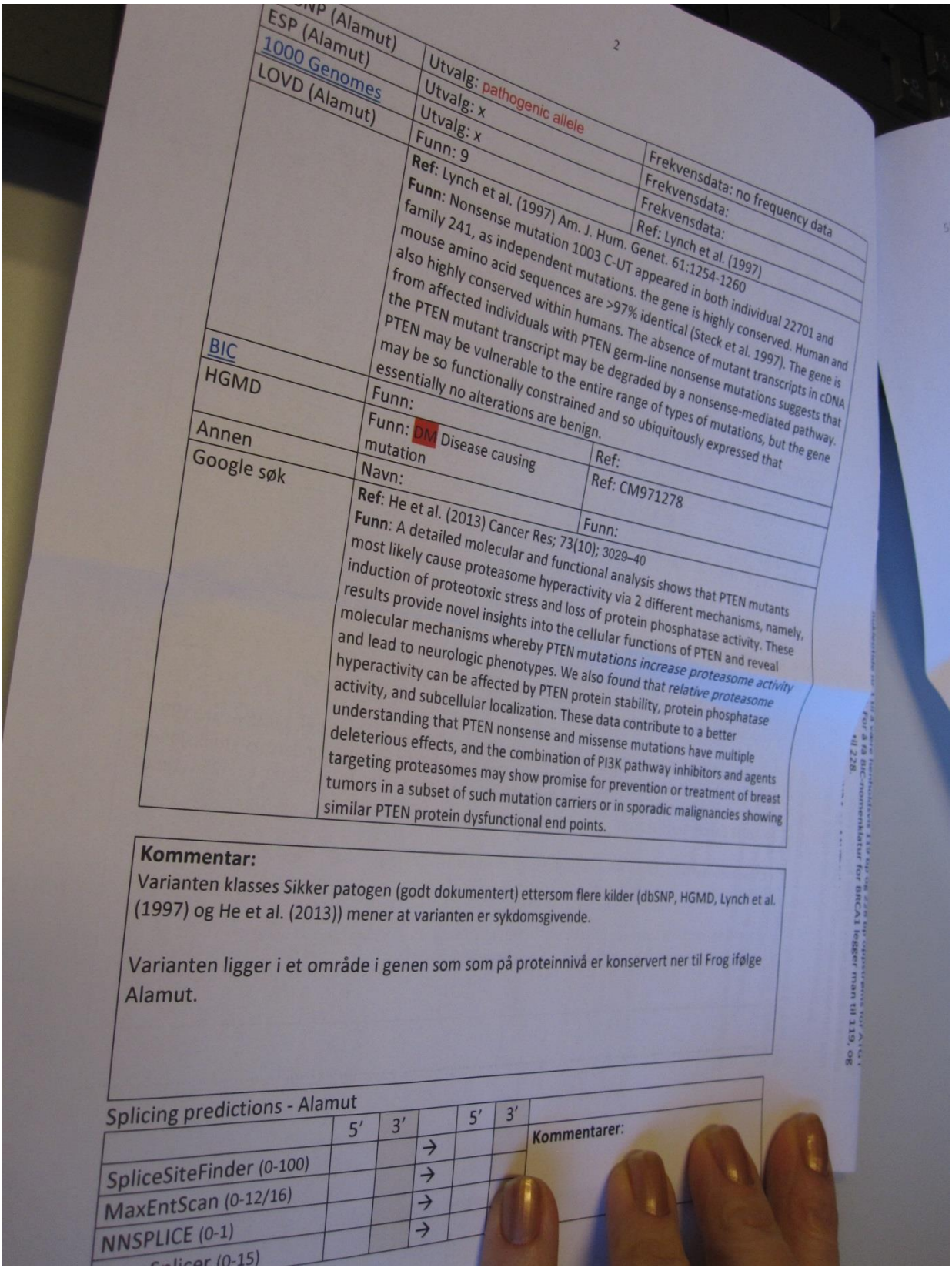



**Figure 13: User developed gene variant and research article classification.**

In picture 1 the answer report is formulated based on the evaluation of two references. The gray colored text is addressing issues that had some elements of uncertainties. The classification system currently used is displayed on the first page. It is described that for class 1 and 5, the research must be deemed as well documented.




**Picture 2: Evaluation sheet.**

In picture 2 the two references are evaluated, described and a comment is formulated to explain the classification of the variant. Findings in the articles are described both by pasting in text from the articles and sometimes rewriting it a bit and simplifying. The text pasted in is

often from the main content of the article and is not restricted to the articles abstract or conclusion. The user states that "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 that place in the article and then read it with their own eyes." The user stated that these sheets are now further developed to also contain type of study, since it is "important to display the type of reference exactly specified." So currently it is "ref" then "study type" and "findings". A red color is used two places to mark that the variant is disease causing.

Through the analysis these issues where identified that should be explored in later user sessions:

- Pasting important parts from the references in the comment-field is important to have for the next evaluator, so that they can find these parts in the text. A functionality where the next evaluator can click on the pasted text and be guided directly to this part in the article might be a valuable functionality.
- There is a need for insuring that the reference schema covers the demand for defining type of study in the reference.
- Since the user would like the list to be sorted based on some sort of criteria. Later user sessions could address what kind of criteria the list should be sorted by.

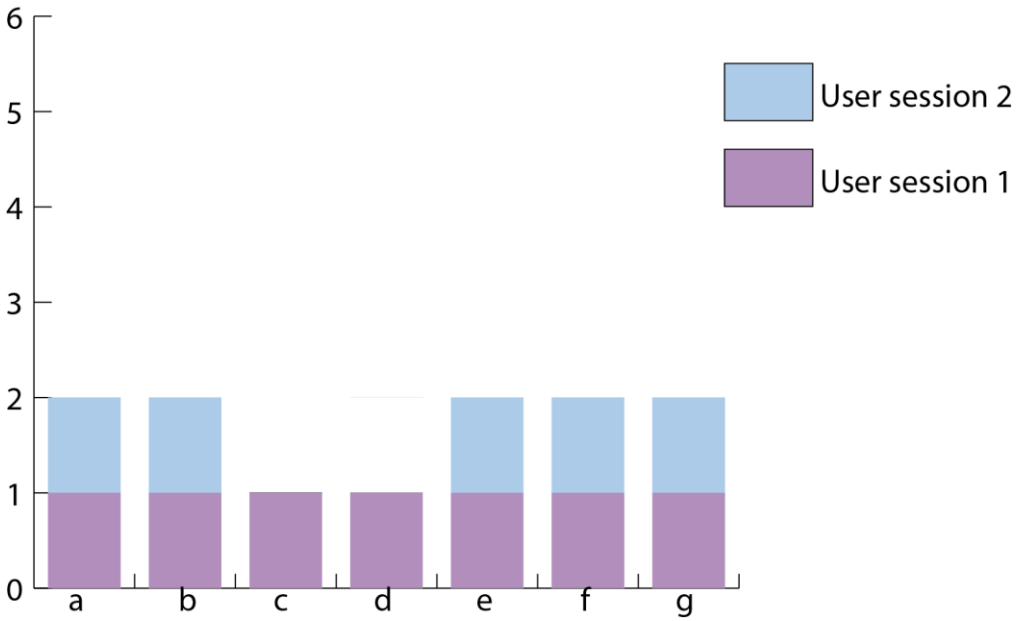
### 8.3.9 The list of themes and issues identified in the first user session

The analysis of the data from the first user session identified 7 specific recurring issues or themes that are displayed in figure 12. This graph will be used to compare these findings in relation to findings in later user sessions. Together these findings will eventually be addressed and validated in a survey. There were also presented other additional findings and a user developed strategy for evaluating research references for gene variant classification. Based on all of these the identified design issues and themes from the initial user session are:

1. The system and especially the comment field must insure that one understands what others are thinking when they make their assessments.
2. The system should provide support for diverse user strategies deployed to evaluate research articles. These strategies are both related to individual user preferences and brought on by the demands from the different variant classification cases.
3. Provide clear options for how to handle assessments that are done with some level of uncertainty. User generated comments linked to some level of uncertainty, could be color marked. To convey the uncertainty to the next user performing the evaluation.
4. Important to display the type of study used in the article exactly specified. To handle that type of study is sometimes difficult to assess. The type of study identified should then be displayed in the final report.
5. In addition to study type, the study's method and material used on the specific variant should be described in the comment field.

6. Write a short summary of the findings in the article related to the current variant.
7. Functionality provided for copying from the articles content and pasting into the comment field. The pasted text should be editable, since users sometimes need to shorten the pasted text.
8. There should be a spellcheck of how the authors are registered in the program when uploading articles. There is an issue in currently used systems that people mistype.
9. Handle future special cases where one reference is linked to and relevant for the classification of many different gene variants.
10. Color coding of evaluated references.
11. Providing the possibility to pause the classification process if needed and go back and retrieve text accidentally lost.
12. Like if one has finished work for the day and wants to continue the next morning. One of the most frustrating user experiences with the comment field would probably be if the user would lose the composed comment without any functionality for retrieving the text.
13. The list of articles should be sorted by some form of ranking criteria.

## 8.4 The second user session




a. Multiple user strategies for deciding which article to evaluate first.
b. Describe study type in list of articles.
c. Color coding of evaluated references.
d. Possible to mark/register issues of uncertainties.
e. Paste quotes from article into comment field together with own comments.
f. Difficult to extract the information in article about the specific variant.
g. Describe study type, material, method in comment field.

**Figure 14: Graph displaying the identified "red thread" issues and themes, from user session 1 and 2.**

### 8.4.1 The most recurring themes and issues identified through user studies

The user in session 2 also states that the type of study is very important and wants to write a summary of the findings in the article. She further explains that in a reference there is often many variants that are studied and it is often difficult to assess which type of study that is done on the specific gene variant that is being evaluated. When asked what is the most challenging with the task of reference evaluation. The user states that the problem is that there might be as many as 20 articles to evaluate and conclusions in the articles might go in different directions, even though they have both based the conclusion on functional studies. Then it is hard to evaluate the validity of the functional studies performed in the articles. The user states that they often prioritize the newest and the articles that have done many different types of analysis on the variant. She further states that there are coming more and more good articles based only on one variant and that uses data from others combined with own studies and makes likelihood models.

### 8.4.2 Uncertainties

The biggest difference between the results from the two user tests are related to how they prefer to handle being unsure of something with an assessment. The second user would handle this by going and discussing it with a locally available colleague. The first user would like to make the assessment, but register somewhere that the assessment is done with some level of uncertainty. Then the next person doing the evaluation sees this and adds their assessment.

There was also a difference in relation to the use of color in the prototype. The first user clearly stated a desire for more color, but the second user did not have any clearly stated color usage preference.

In the prototype the search functionality that generates searches in Google Scholar covering all the different variant names used on the gene variant is highly valued by the users. Based on the user test it should also be added regular Google. The user explains that the lab doctor gets the references after the two first users doing the evaluation are finished.

Through the analysis these issues where identified that should be explored in later user sessions:

- The persons that have done prior evaluations are now locally available for verbal discussions, but if the program is going to be used at other hospitals. The system has to support global collaboration where the teams doing the evaluations are geographically distributed.
- There is a need for insuring that the reference schema covers the demand for defining type of study in the reference. Another possibility is to add a field for the user to type in type of study.

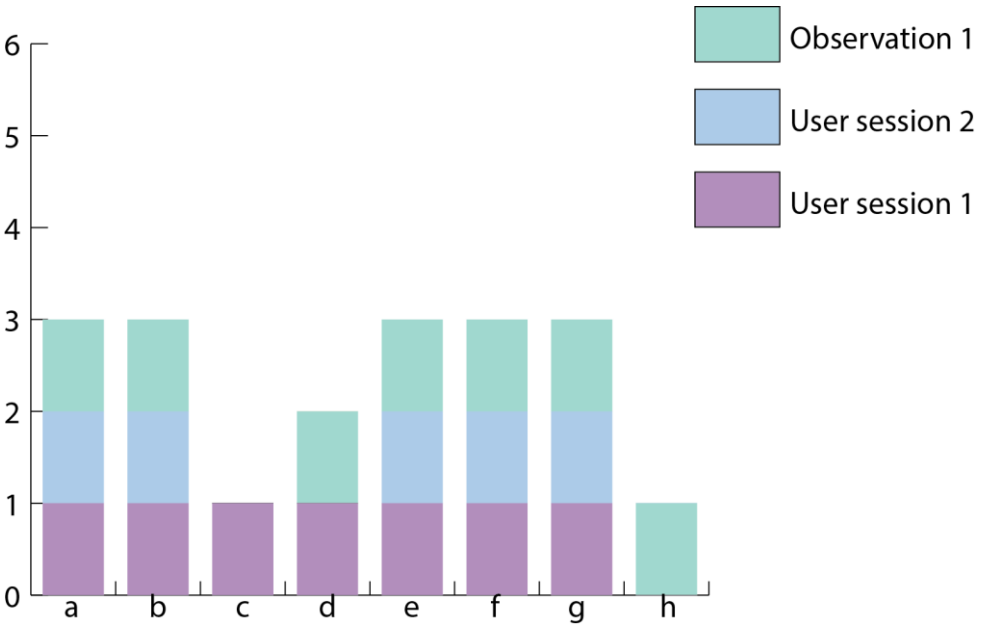
- Since the user states that they often prioritize articles that are the newest and that contain many different analysis of the variant. It might be valuable to sort the list of articles by placing such articles at the top of the list. The articles focusing on others findings together with own findings and using likelihood models might also be relevant to prioritize.
- Since the lab doctor gets the references, maybe they should have access to the program and if they evaluate the references it should be addressed if and how their assessments of the articles are stored in the system.

### 8.4.3 The list of themes and issues identified in the second user session

The design themes and issues identified in session 1 were also relevant in user session 2. In addition, there was new design issues identified.

14. The user wants the articles marked as not relevant by the first person doing the evaluation to still be assessable for the second person doing the evaluation. Since there might be a disagreement upon the articles relevance between the first and second person doing the evaluation.
15. Important to display who that has done prior evaluations. Since the user will contact them if they need to discuss a prior assessment.

# 8.5 Observation 1




a. Multiple user strategies for deciding which article to evaluate first.
b. Describe study type in list of articles.
c. Color coding of evaluated references.
e. Possible to mark/register issues of uncertainties.
f. Paste quotes from article into comment field together with own comments.
g. Difficult to extract the information in article about the specific variant.
h. Describe study type, material, method in comment field.
i. Searches PDF to see if and how many times the variant is mentioned.

**Figure 15: Graph displaying the identified "red thread" issues and themes, from user session 1, 2 and observation 1.**

## 8.5.1 The most recurring themes and issues identified through user studies

Most of the themes identified in the prior user sessions were also identified in the data from this observation and the semi structured interview performed after the observation. There were no color usage preferences identified.

It is observed that the evaluation is in some ways based on a kind of gut feeling. It is stated that if one sees any mistakes, one wonders what other mistakes they made and the results are trusted less. In observation of the evaluation of one reference one such mistake is identified. It is an article where they are inconsistent in the name used for the variant; they use two different names for a variant. This is perceived as a careless mistake and is noted in the evaluation form, but the reference is still used in the classification of the variant together with other articles with similar findings.

When evaluating the article, they are looking for what the study is about. If it is only in silico / likelihood-ratio, then the results indicate something, but their conclusion will not be made based only on that. There are strict criteria that they must find articles with functional studies to make a classification.

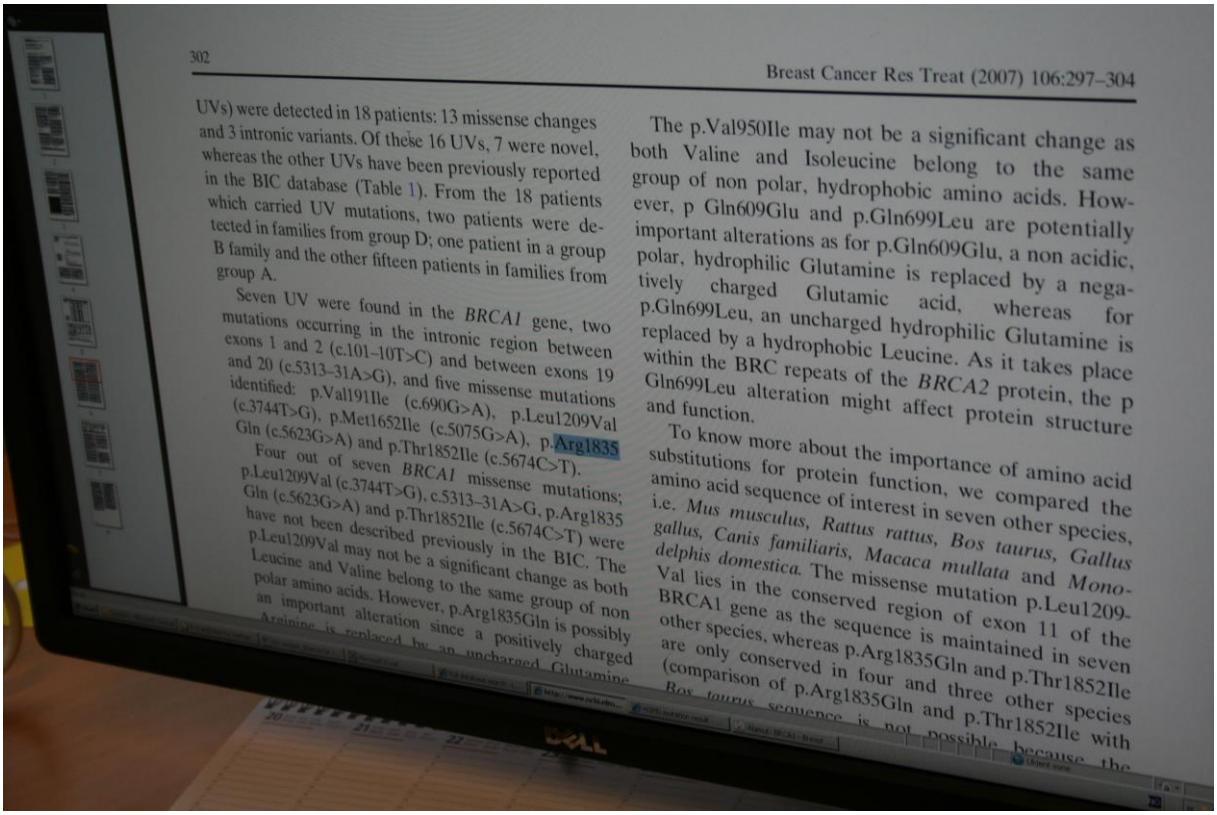



**Figure 16: User generated variant key word search of PDF research article.**

The main finding from the observation related to the evaluation of references, was that when the user opens a reference the user searches through the reference to see if the variant is mentioned. When doing this search in the PDF of the article, the users have to types in many different names used on the variant before finding the one used. The user states that they search up the variant to get an initial impression of the article. They use it as a strategy to decide which article to evaluate first from the list of articles that they are going to evaluate. There is a problem that many of the articles for evaluation do not mention the variant, or only in Supplementary information. Since they only need to find two articles of high quality with the same conclusion to classify the gene variant they can stop evaluating the list when two such articles are found. If the variant is mentioned many times the reference is perceived as more interesting. They also look for the newest and articles containing many functional studies, this influences which article they evaluate first. The users find the references through the Alamut program that generates Google searches covering all the names used on the variant. They state that they value this functionality highly.

Through the analysis these issues where identified that should be explored in later user sessions:

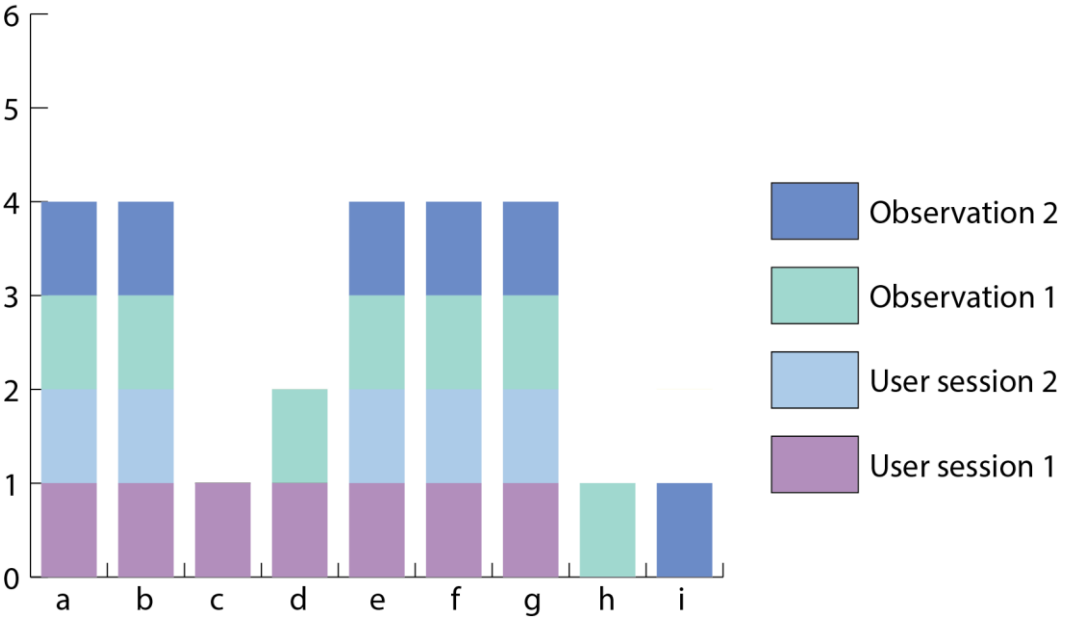
- In the sorting of the list the newest and the ones containing functional studies could be prioritized together with the ones where the variant is mentioned the most.
- Since how many times the variant is mentioned in the article influences which articles are evaluated first. The system could present a word count next to each article in the list of references.
- When searching the PDF for the name used on the variant the user have to try the different names before the right is found. The system could automatically search the PDF and display which name that is used.

### 8.5.2 The list of themes and issues identified in the first observation

The design themes and issues identified in session 1 and 2 were also relevant in the observation. In addition, there was new design issues identified.

16. It should be registered in the evaluation if an article contains careless mistakes.
17. Insure that the all the articles presented in the list of references to evaluate mentions the variant.

# 8.6 Observation 2




a. Multiple user strategies for deciding which article to evaluate first.
b. Describe study type in list of articles.
c. Color coding of evaluated references.
d. Possible to mark/register issues of uncertainties.
e. Paste quotes from article into comment field together with own comments.
f. Difficult to extract the information in article about the specific variant.
g. Describe study type, material, method in comment field.
h. Searches PDF to see if and how many times the variant is mentioned.
i. Possible to store other formats than PDF.

**Figure 17: Graph displaying the identified "red thread" issues and themes across performed user sessions.**

## 8.6.1 The most recurring themes and issues identified through user studies

Most of the themes identified in the prior user sessions were also identified in the data from this observation and the semi structured interview performed after the observation. There were no color usage preferences identified or concerns regarding issues of uncertainties.

There are two laboratory engineers working with the classification of genetic variants regarding hereditary heart disease. The user states that if they are unsure of an evaluation they discuss it and makes the assessment in collaboration. They have contact with colleagues in Trondheim and Bergen and the Rikshospitalet University Hospital. The user states that for example, those from Bergen can call and ask is the classification of the variant is certain? They can answer yes, but not send them any data. This contact is through phone calls or e-mails.

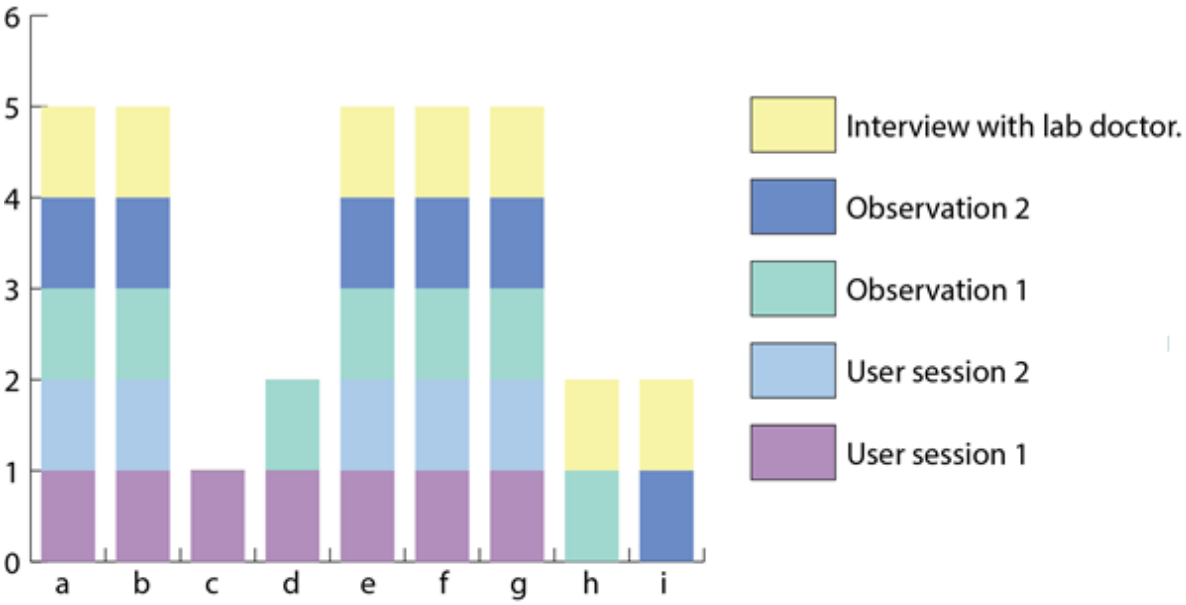
The user states that there was a gene variant that the user did not find any articles about. But in Google he found a poster from a conference that had addressed the variant that was valuable for the classification process.

### 8.6.2 The list of themes and issues identified in the Second observation

Most of the prior identified design themes and issues were also relevant for the observation. In addition, there was a new design issues identified:

18. Possibility to store other data than PDF references like conference posters if these were important for the classification.

## 8.7 Interview with laboratory doctor




a. Multiple user strategies for deciding which article to evaluate first.
b. Describe study type in list of articles.
c. Color coding of evaluated references.
d. Possible to mark/register issues of uncertainties.
e. Paste quotes from article into comment field together with own comments.
f. Difficult to extract the information in article about the specific variant.
g. Describe study type, material, method in comment field.
h. Searches PDF to see if and how many times the variant is mentioned.
i. Possible to store other formats than PDF.

**Figure 18: Graph displaying the identified "red thread" issues and themes across performed user sessions.**

### 8.7.1 The most recurring themes and issues identified through user studies

Most of the themes identified in the prior user sessions were also identified in the data from this interview. There were no color usage preferences identified or concerns regarding issues of uncertainties. The prior identified difficulty to extract the information about the specific variant was a recurring issue in this interview.

When asked what is the most challenging with the task of reference evaluation. The user states that the problem is related to the need to know a lot about each variant, in order to judge if it has an impact on disease development. The different study's performed and material used can be good, but then it is difficult to extract the individual components. What is used on the specific variant? If it is just prediction then they are a bit skeptical to the results. Especially the description of study type, material and method is important to identify. Like how many family members are included in the study? The lab doctor often evaluates the material and method used in the study. She describes the focus of this evaluation as "if it is found in a person, who is this person? Are they sick they are healthy? Is there a control population? I will be very focused at identifying this. Especially who are included? At a very detailed level what was the inclusion criterion? Several types of cancer?  First, in general for all variants then specified for the current variant. How many relatives and tissue samples were used for this particular variant?" (lab doctor)

She states that to solve this issue the article could have been written so that it provides more detailed information. She states "I think they certainly have the information. It is omitted because of other reasons like constraints of max word etc. Sometimes you can be lucky and find supplementary online tables with useful information."

One of the most interesting findings in the interview is that the lab doctor has the same strategy as the users in observation 1. She also searches in the PDF to see how many times the variant is mentioned. After being specifically asked if she uses this search strategy in the reference. She states that "the first thing I want to know is what the article says about the specific variant" (lab doctor). She further states that to examine that "I also think it's really smart to search for the variant name in the article. The name used may be different, but once you've figured out what they use. If it is mentioned only once in a schema that gives me less than if it is actually mentioned in the text so I always do the search to get a visual impression." She further states "If I see that here it says a lot about the variant I'm interested in, then I read the article more closely" (lab doctor).

### 8.7.2 Evaluation of research articles

The laboratory doctor gets the evaluation form from the laboratory engineers together with the references. She states that "they have become very good at filling out these forms so that when they have obtained frequency data and used prediction programs to see if there is a new splice site. Then I trust what they have done." She says that she does not have any more

knowledge then them on these issues. Where she often contributes is in the evaluation of the articles. She states “where I often can come in and do a reasonable job is with the literature search and it may well be that they have already found the articles, or I find more by myself. But then read them in a bit more detail.” She can have a deeper understanding of if the effect of the found variant is small or big and the consequences for the patient based on the effect of the gene mutation. She states “the difference between an association which may have a small effect and a real cause of something.” She uses Alamut to search for articles and states that she likes the Alamut search function for finding articles. Sometimes she also uses Google and PubMed.

Articles may be relevant for more variants and sometimes articles are recurring for more cancer genes that are addressed in the reference. These are not noted down, but are remembered and sometimes they go back to older evaluation forms and retrieve the article. The evaluation of a reference in regard of one variant cannot be reused directly, but it can be valuable to evaluate for the current variant. The system could keep the articles available for all the relevant variants.

### 8.7.3 Medical language

Another area where she contributes is the understanding of medical jargon (like Medical Latin), that a bioengineer cannot be expected to understand, but laboratory doctors understands. She states “this may for example be descriptions of skin changes or Cowden syndrome.”

The laboratory doctor states that “we have much to gain from increasing the efficiency of the analysis process.” She exemplifies by describing how the waiting time for receiving the results from a genetic testing for BRCA1 and BRCA2 (Breast and ovarian cancer) is drastically reduced. Recently in the autumn of 2013 they have introduced a rapid test for patients waiting for operations that take approximately one week. A couple of years back it took one year from the requisition was sent from the patients doctor, then if approved it took one more year before the genetic testing were done and the results were received.

Through the analysis these issues where identified that should be explored in later user sessions:

- Since the lab doctor used the reference search function in Alamut, maybe the lab doctor would be interested in using this function in the work bench
- Since the lab doctor evaluates references and ads more references if needed should this be registered in the workbench and if so how?
- Maybe there should be added fields for the study method and material. That the lab doctor’s evaluation of these should be added.
- If there are two or more genes that in combination make a pathogenic effect, how should the system handle this? The references should then be linked to two gene variants and the evaluation should then be on the combination of the two variants. And

if the users discover such a combination effect of two variants how should they register this in the system to get the possibility to evaluate the two or more variants in combination. The system needs to be flexible enough to handle such special cases.

- what that is included in the text in the comment field by the evaluators is described as individually different

### 8.7.4 The list of themes and issues identified in the interview with a lab doctor

This session also identified a need to store other data than PDF references, more specifically supplementary research data found online. One of the most recurring issues was the prior identified need and difficulty to identify and register study type, method and material.

19. Support easy access to articles that are recurring for many cancer genes that are addressed in the reference.

# 8.8 Analysis and discussion of the results from the survey

As mentioned in the methods chapter a survey was used to validate findings and investigate them further. The results from the survey (Attachment C) were analyzed with qualitative content analysis. This is a presentation of the results from the analysis.

## 8.8.1 Multiple user strategies for deciding which article to evaluate first

After negotiation with coder 1, we came up with 6 categories that were agreed upon. They are listed in the table 3 and displayed in figure 19

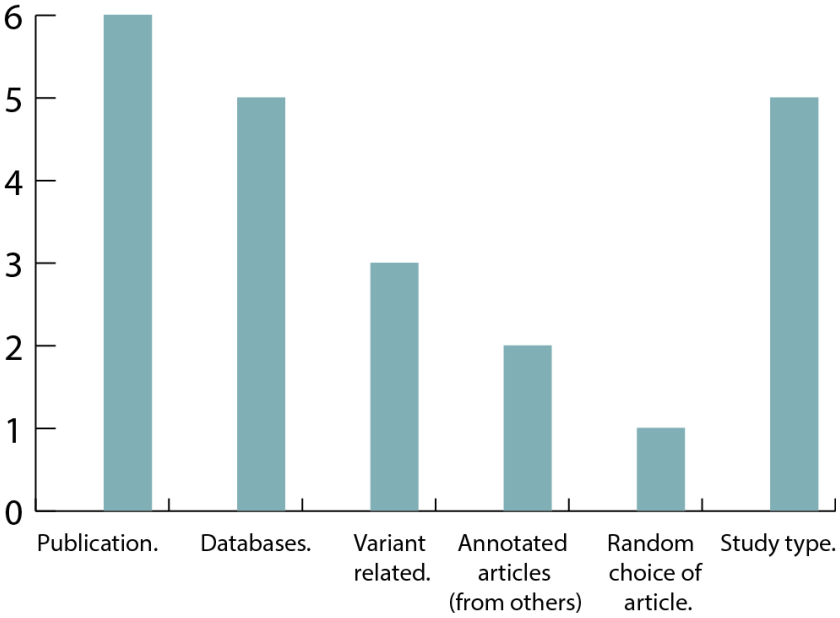



**Figure 19: Which articles that are evaluated first.**

| Question:<br>If you find more references for the current variant, how do you choose which one to evaluate first? | |
|---|---|
| Categories | Details |
| 1) Publication. | Newest, tittle, author(s), journal, abstract |
| 2) Databases. | Google, HGMD, LOVD, Alamut. |
| 3) Variant related. | Relevance for variant and Variant mentioned in article. |
| 4) Annotated articles (by others). | Descriptions from others. |
| 5) Random choice of article. | |
| 6) Study type. | Functional study, segregation analysis, family information. |

**Table 3: The results from the first question in the survey, regarding which articles that are evaluated first.**

Many of the responders mentioned different elements that effected which article they choose to evaluate first. Some also stated one specific strategy they preferred above others. These are listed and counted in the table 4.

| Categories | Users | Total |
|---|---|---|
| Start with the newest first. | #10 | 1 |
| Based on a description from others. | #1, #3 | 2 |
| Randomly | #7 | 1 |
| A combination of different elements from the article | #2,#4,#6 | 3 |
| Starts with articles containing functional studies first. | #5, #8, #9 | 3 |
| Did not mention one specific strategy they used. | #11 | 1 |

**Table 4: Strategies to select the first article to evaluate.**

All the different elements that users stated as important factors are listed in table 5.

| Categories | Users | Total |
|---|---|---|
| Author(s) | #14, #6 | 2 |
| Journal | #2, #4 | 2 |
| Abstract | #9, #10 | 2 |
| Age(newest) | #2,#3, #4,#6,#10 | 5 |
| Tittle | #4, #6 | 2 |
| Priority in google search | #5, #4 | 2 |
| Variant mentioned in article | #6,#9,#11 | 3 |
| Articles containing functional studies | #2,#5,#6,#8,#9,#10 | 6 |
| Family information | #2,#6 | 2 |
| gene frequency data | #6 | 1 |
| Insilco | #10 | 1 |
| Observation studies | #10 | 1 |
| Segregation analysis | #2,#6,#9 | 3 |
| Study based on predictions | #6 | 1 |

**Table 5: All the identified elements influencing which articles that are evaluated first.**

Some of the strategies presented in table 5 were not always possible. For the user to start with articles containing functional studies, the information about which articles that contains functional studies has to be available. Similar there has to be a description from others that has done a prior evaluation of the article on order for this strategy to be possible. If this

information were present these strategies were decisive for their choice of first article to evaluate.

These results further validate the need for the system to support multiple user strategies. Specific information must be available for the user when the user makes the decision of which article to evaluate first. The information needed to deploy the strategies in table 5 is crucial. Namely the publication year, a description from prior evaluations and that it is specified if the articles conclusion regarding the specific gene variant is based on functional studies. The articles relevance for the specific variant can be indicated in some way by presenting which studies that is used to examine the specific variant.

### 8.8.2 Evaluation form for references

Regarding the first question about the evaluation form for references the negotiation with coder 1 resulted in 5 categories. They are listed in the table 6 and displayed in figure 20.

| Question: We have chosen to include segregation, protein, RNA, coverage and age of the reference as categories that may affect whether a reference can be evaluated to be of high quality. Does this make sense? Is there anything else you think should be included? | | |
|---|---|---|
| Categories | Users | Total |
| Yes. | #11, #1 | 2 |
| Yes, with other suggestions. | #2,#3,#7,#9, #10 | 5 |
| No. | | 0 |
| Only remarks. | #4,#6,#8 | 3 |
| No answer. | #5 | 1 |

**Table 6: Segregation, protein, RNA, coverage and age of the reference as indicators of reference quality. Should more elements be included?**

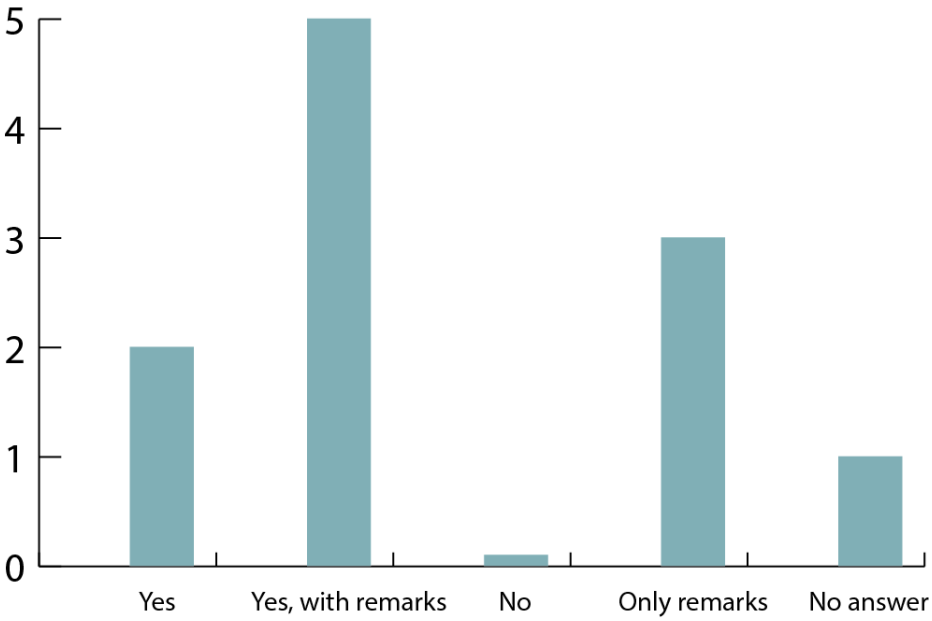



**Figure 20: Segregation, protein, RNA, coverage and age of the reference as indicators of reference quality. Should more elements be included?**

There was a problem related to information needed to fill out the form that was missing in many articles. Examples are gene coverage and segregation data. The LOD score is also rarely mentioned in the article. One user suggests handling these issues by adding a button for "not supplied".

One responder states that it is perceived as confusing how in the form one is to answer "Does reference support pathogenicity?" not only by the options yes and no but also with "VUS". The user suggests removing the "VUS" option and adding a question, "Does reference support neutrality?" With the options yes and no, where no is used for VUS variants.

To insure the validity of the results from the form, there should be specified requirements for the data. For example two responders address the importance that the segregation data is based on multiple families. One family is not sufficient to classify the variant as pathogenic. Four responders' address a need for specifying that the information filled in the form needs to stem from functional studies and not prediction studies. Especially protein and RNA information needs to be derived from functional studies, to have importance for the quality of the article. One user suggests handling this by clearly stating in the program that the data should only be derived from functional studies. Two users also state that they in addition to functional studies find studies using multifactorial likelihood analysis to be convincing and they should be included. One responder stated that the publishing journal and if the article is perceived as well written is of importance. If careless mistakes are found the article loses some credibility and this should be registered.

When asked if there are other things that should be included in the form, the responders mentioned different elements. There is identified a need for the system to handle differences in the variants being classified. Like when classifying genes with dominant inheritance, users

also identify if the variant is found together with another disease causing variant. One responder stated that gene coverage is different in some genes (like the gene CDK4) that are sequenced by using single exon sequencing. A responder would like to add if the variant is described in the article to cause changes in splicing. Another responder states that for the Lynch genes it is interesting with data from studies using immunohistochemistry and microsatellite instability.

The next question was also regarding the evaluation form for references. The users were asked whether the category "segregation" should be more differentiated. The results are presented in figure 21. One user would like to add the number of people tested. How many with/without phenotype that is positive/negative for the variant. A responder states "the number of unrelated families, number of generations, the number of people in the family could be interesting" (#4). Another user states "in cases where there is a probability of pathogenicity based on segregation, it will usually be listed as a likelihood ratio, this might be valuable to fill in" (#10).

| Question: Should the category "segregation" be more differentiated? Like a possibility to fill out a LOD score, or number of generations? | | |
|---|---|---|
| Categories | Users | Total |
| Yes. | #2,#7,#8,#9 | 4 |
| Unsure. | #1,#3,#4,#6,#10 | 5 |
| No | #11 | 1 |
| No answer. | #5 | 1 |

**Table 7: Should the category "segregation" be more differentiated.**

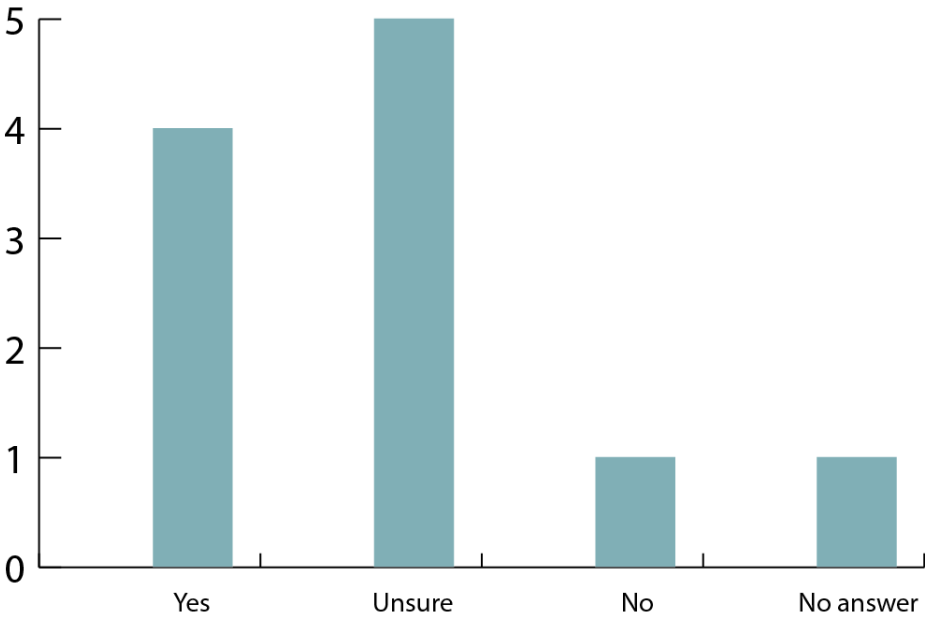



**Figure 21: Should the category "segregation" be more differentiated?**

### 8.8.3 Comment field

All of the users answered yes to the question; "Do you think it is useful to paste important quotes from the article into the comment field?"

They were also asked to specify how they selected the specific quotes that they copied and pasted into the comment field. Most of the users stated that the quotes that were selected were relevant for the specific variant and had meaning for the classification of this variant. The following three categories were repeatedly stated by the users as important quotes:

- Concluding quotes regarding the specific variant.
- The kind of data that was used to examine the specific variant.
- Description of the method used in the study of the variant (e.g. functional study/Insilco).

Based on the answers from question 8 and 9, the elements the users find beneficial to include in the comments is presented in table 8.

| Categories | Users | Total |
|---|---|---|
| The articles conclusion regarding the variant | #1, #2, #3, #4, #5, #6, #10, #11 | 8 |
| frequency data | #1, #3 | 2 |
| Information about patients in article (sick / healthy) | #1, #3, #4 | 3 |
| splicing results | #1 | 1 |
| Description of method (Functional study/ What kind of Insilco methods used) | #1, #4 #9 | 3 |
| Functional study | #1, #3, #6, #9 | 4 |
| Family information | #1, #4 | 2 |
| If the article is well written. | #1 | 1 |
| A summary of the article | #1 | 1 |
| Personal opinions about the article | #1, #3, #7, #10 | 4 |
| Information about date and name of people that has done prior evaluations. | #2, #4 | 2 |
| Other articles referred to in the article with same/different conclusion. | #5 | 1 |

**Table 8: The elements the users find beneficial to include in the comments.**

User #5 requested a possibility to write the comment in Norwegian.

It is identified a need for insuring that the data written by the prior evaluator is perceived correctly. The comments might be based on the article statements combined with subjective statements and these should not be uncritically blended, to insure no confusion between the two. One user asks the question: “Sometimes quotations, sometimes shorter summary / descriptions of the findings. How do you distinguish quotes and summaries?” There are also 4 users that state that they want to write personal opinions in the comment field and their own assessment of the articles quality and findings.

### 8.8.4 Information shown in reference overview table

There is identified a need for increasing the information shown in the reference overview table. Two responders want more information about the previous evaluations. Specifically what are they? And which variants were they done for? Three of the respondents wants it to be displayed what kind of study the reference has used to investigate the current variant. One user states "I would like it to appear in a separate column if it has been done functional studies" (#9). Another user (#4) wants it to be displayed if the reference is a primary literature report, functional characterization or phenotype report. And states further "If you have two phenotypic articles and one or two functional pointing in the same direction, then it is pretty strong" (#4). One responder states "I think it would be helpful to have the score and the comment in separate columns" (#5).

It might be valuable if it is displayed what type of study the reference is based on and if the evidence in the article is pathogenic/neutral or VUS. That way the user can quickly see if there are functional references and phenotypic references that both indicate the same classification. This might be helpful for the next person to evaluate the references as well. They might get an impression of the strength of the evidence used in the previous evaluation immediately.

### 8.8.5 Word count from article

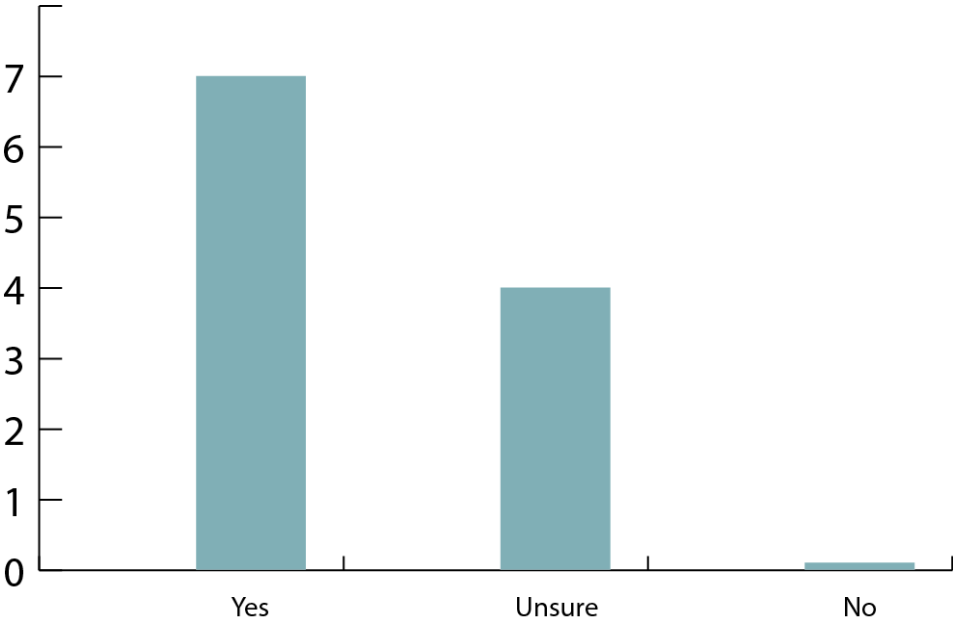



**Figure 22: The results related to the word count functionality.**

| Question: Will this word count be useful to you (and why)? Specified: Answers related to counting the different variant names used for current variant. | | |
|---|---|---|
| Categories | Users | Total |
| Unsure | #4,#6,#7,#8 | 4 |
| Yes. | #1, #2, #3, #5, #9, #10, #11. | 7 |

**Table 9: The word count.**

| Question: Will this word count be useful to you (and why)? Specified: Answers related counting the words “RNA”, “Protein” and “Segregation”. | | |
|---|---|---|
| Categories | Users | Total |
| No | #2,#7,#8,#9 | 4 |
| No answer | #1, #3, #4, #5, #6, #10, #11 | 7 |

**Table 10: The word count.**

When asked if they find it useful to be provided with a word count showing how many times the variant is mentioned in the article. None of the responders answered no, it is not useful. Seven of the responders answered yes without hesitations. There were four responders that were unsure regarding how useful it is in regard of being an indication of the articles quality. There were many responders that would use it to decide which article to evaluate first, since the articles with high “word counts” would have focused a lot on the variant. In addition to frequency some unsure responders also mentioned position of the variant name in the article as important. They stated that if the variant is just listed in a table or in supplementary data it is not helpful for the classification of this variant.

When asked if there “are any other keywords you think are relevant to search in this way?” One user stated “frequency and other words related to that” (#3). Two users specified that it is important to include all possible different names describing the specific variant.

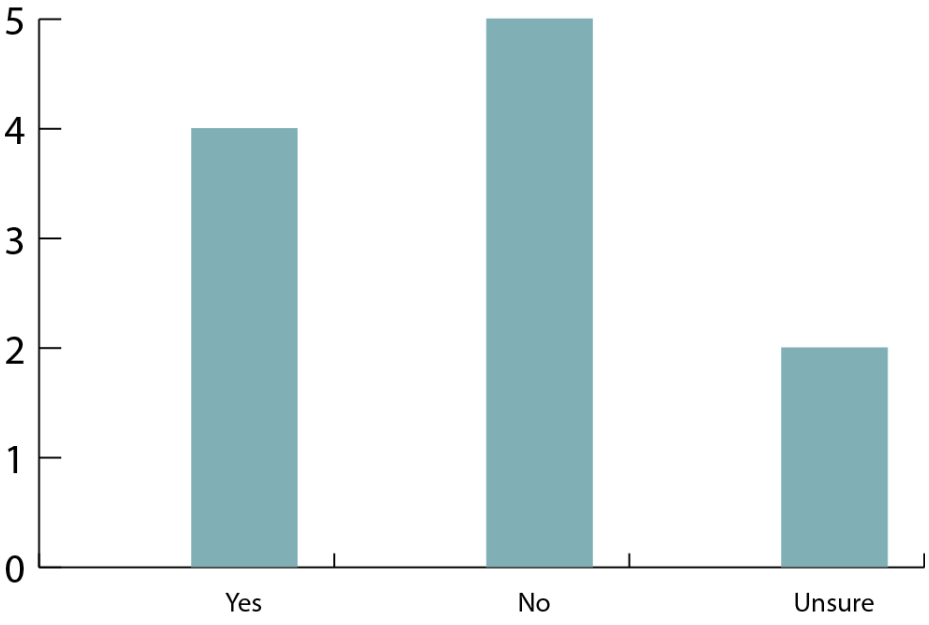



**Table 11: Results related to comparing word counts between articles.**

| Question: In cases where there are several references about the same variant: is it desirable to compare the word count of these? How would you use this? | | |
|---|---|---|
| Categories | Users | Total |
| Yes. | #3,#6,#8, #11 | 4 |
| No | #1, #2, #5, #7. #9 | 5 |
| Unsure | #4, #10 | 2 |

**Table 12: Compare word count between references.**

When asked specifically if it was desirable to have functionality displaying comparisons of word counts in graphs. Four responders stated that they would use the variant frequency word count to select which reference to evaluate first. Four other responders said that they were not interested in variant frequency comparison between articles. One user states "This does not necessarily say anything about the article quality. A variant may be mentioned only two times (e.g. the results and a table), while the rest of the article describes the method very well. Is the method good and there is evidence that the variant is assessed by this method may be enough to use the article for the classification."

### 8.8.6 Color coding of evaluated references

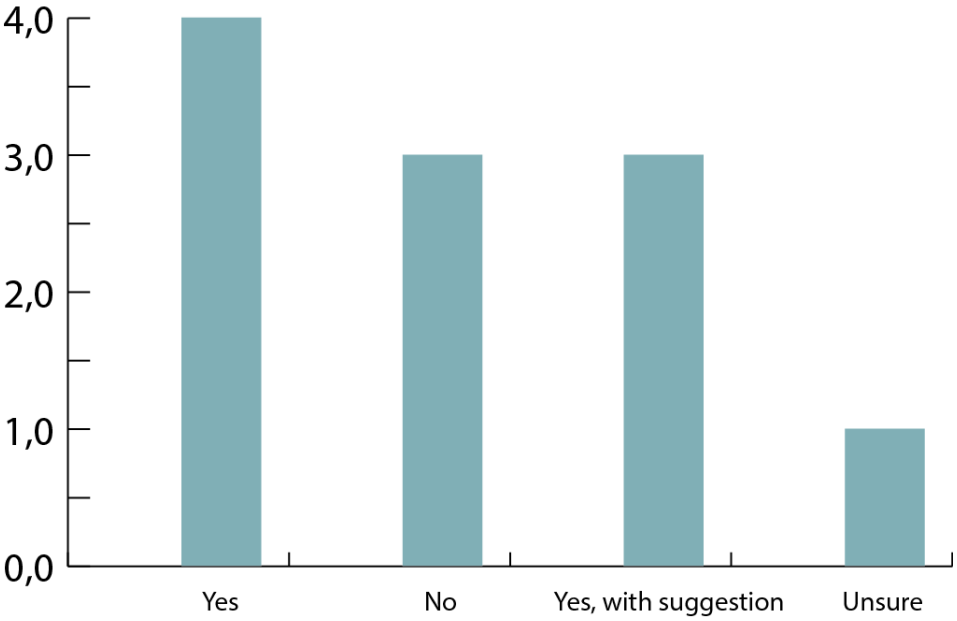



**Figure 23: Positive/negative to color coding evaluated references.**

| Question: What do you think about the highlighting of references like presented in the picture, and how do you think it should be done? | | |
|---|---|---|
| Categories | Users | Total |
| Yes. | #2, #6, #8, #11 | 4 |
| No. | #5, #7, #10 | 3 |
| Yes with suggestion | #1,#3, #4 | 3 |
| unsure | #9 | 1 |

**Table 13: Positive/negative to color coding evaluated references.**

When asked about their opinions about using color to highlight prior evaluated references. Seven responders were positive to the color coding. Three of these had additional suggestions. A responder suggested adding a "field/route for pathogen/VUS/neutral. That is colored respectively red/yellow/green." Another responder stated "I would like to see a clear color marking on the header row (header, light blue/light green/light yellow/maybe dark gray/black with white text) so that one can easily identify this field" (# 4).

One of the responders was unsure about color coding, but suggested red for pathogen and green for neutral variants. She warned "I think that color coding can easily make you become uncritical, drawing conclusions too quickly without really considering the content of the reference. Color coding must be organized so that there are few colors so that one at all times knows what for example purple and yellow means and it must be used with caution" (# 1).

It was identified that even though seven of 11 responders were positive to color coding. There was also a general preference for displaying issues with written information. One responder stated "it is better with written conclusions/summaries" (# 10). A responder stated "Instead of color coding have a column named "comment" where the person that has evaluated the article adds a comment stating if the reference is particularly good/bad and if it contains functional studies or in silico analysis of the variant" (# 9). Another responder stated "I think you should limit the number of colors, instead you can highlight the score. One can also have yes/no (score "X") on a line by itself, and that the sentence after the score starts on a new line or another column, this may make score more distinct." (# 5).

### 8.8.7 Design and overall structure of the reference evaluation

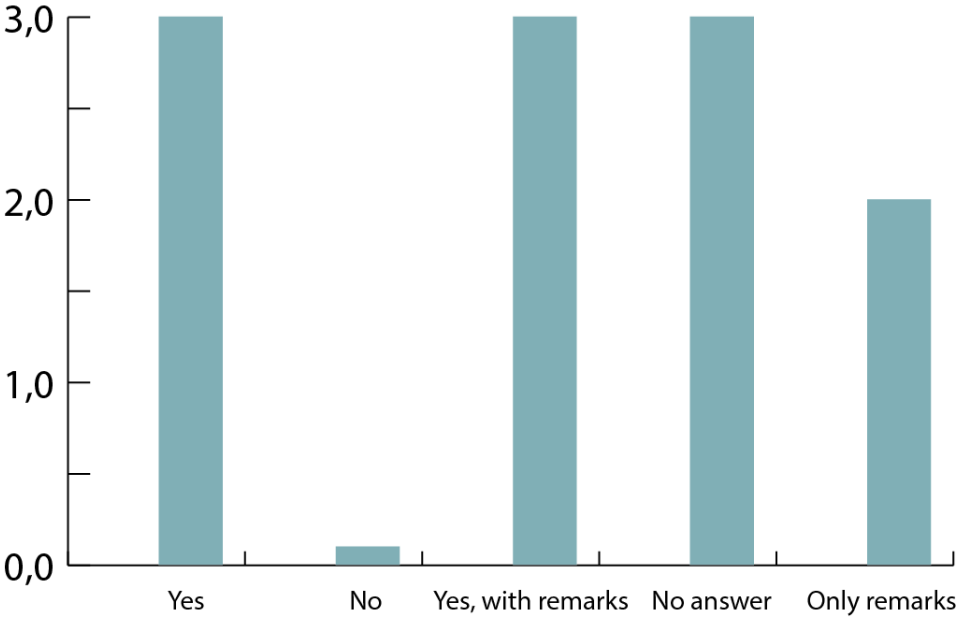



**Figure 24: User satisfaction with program.**

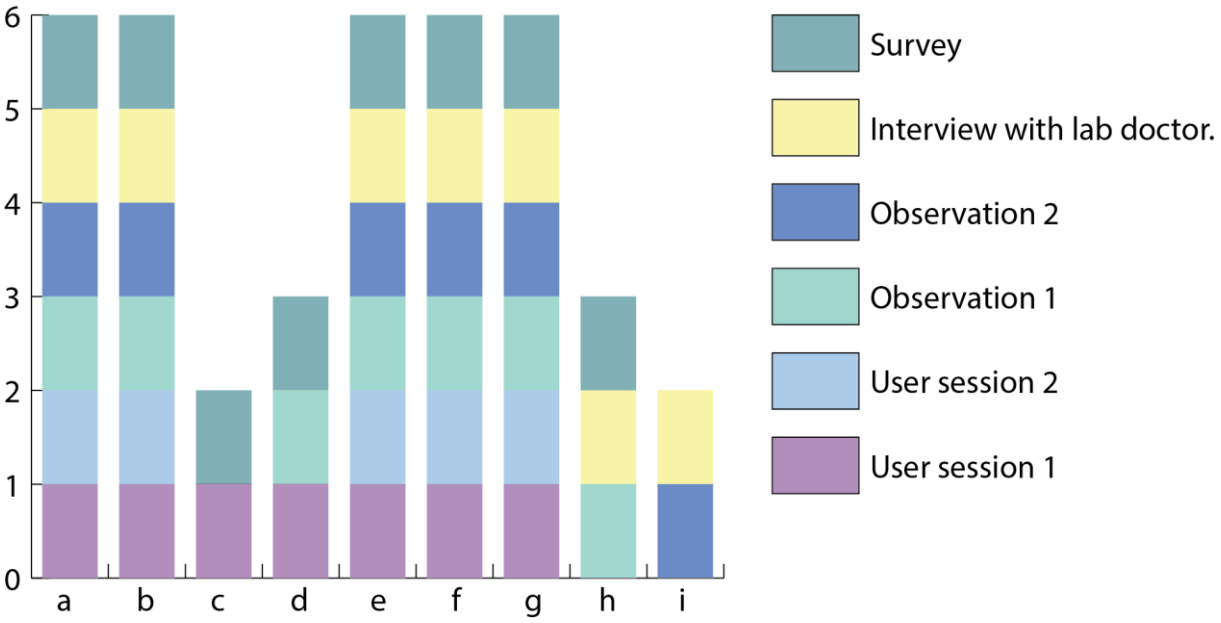
| Question: Does the program as it is described in this survey cover how you currently evaluate a reference? If no, what is not covered? | | |
|---|---|---|
| Categories | Users | Total |
| Yes. | #3 ,#6, #7 | 3 |
| No. | | 0 |
| Yes with remarks | #1, #9, #11 | 3 |
| No answer | #5, #8, #10 | 3 |
| Only remarks | #2, #4 | 2 |

**Table 14: User satisfaction with program.**

When asked "does the program as it is described in this survey cover how you currently evaluate a reference? If no, what is not covered?" No responders answered no. There were 3 responders that did not fill in an answer. Two only had remarks. Three responders wrote yes without any remarks. There were three responders that wrote yes, but added remarks. One respondent stated "some variants will fall in between yes/no. So it is important to differentiate the answer when needed." Other remarks were:

- It is important that one can conclude that the article was good in spite of the variant being classified as a sure class 3.
- It should be displayed much clearer if the article is based on Insilco or functional studies.
- Articles concerning segregation are not emphasized more than articles based on protein or RNA. It is the combination of all of these together that counts.
- The variants are slightly differently evaluated in different genes, such as genes that have different isoforms.
- Maybe need for one Re-evaluate button?

### 8.8.8 Prior identified themes and issues validated in the survey




a. Multiple user strategies for deciding which article to evaluate first.
b. Describe study type in list of articles.
c. Color coding of evaluated references.
d. Possible to mark/register issues of uncertainties.
e. Paste quotes from article into comment field together with own comments.
f. Difficult to extract the information in article about the specific variant.
g. Describe study type, material, method in comment field.
h. Searches PDF to see if and how many times the variant is mentioned.
i. Possible to store other formats than PDF.

**Figure 25: Graph displaying the identified "red thread" issues and themes across all user sessions.**

#### 8.8.8.1 The most recurring themes and issues identified through user studies

There is a large amount of findings in user sessions prior to the survey. When planning the survey and testing it in a pilot test, it was evident that only some selected findings could be specifically addressed in the survey, since the survey could not take too long to fill out. The issues addressed with specific questions were a, b, c, d, e, g, and h. The issues not addressed directly were d and f. Through the analysis of the data it was also identified that f. was a recurring issue in the data. The findings validated all the prior identified findings that were addressed.

### 8.8.8.2 The list of design issues extracted from the survey

The survey identified many of the same themes and issues as user sessions and observations. However, there were a few additional issues that became visible through the survey.

20. Accommodate a user preference for displaying issues with written information
21. Be critical towards how the structure of the comment field affects the comment formulated by the user. To insure that the user does not uncritically blend the article statements with subjective statements. To provide quality and make sure that there are no confusion when the next person is doing the evaluation.
22. Provide an option to select information not supplied, when information is missing in the articles.
23. To insure the validity of the various data entered into the evaluation, there should be specified requirements for the data. More specifically in regards of study type, method and material used.
24. There is identified a need for the system to handle demands brought on by differences in the variants being classified. Some type of variants is evaluated differently.
25. It should be displayed what type of study the reference is based on and if the evidence in the article is pathogenic/neutral or VUS. Sometimes it is the combination of references containing different studies that is decisive for the classification.

## 8.9 The total list of themes and issues identified through user studies

A unified presentation of all the themes and issues identified through user studies:

1. The system and especially the comment field must insure that one understands what others are thinking when they make their assessments.
2. The system should provide support for diverse user strategies deployed to evaluate research articles. These strategies are both related to individual user preferences and brought on by the demands from the different variant classification cases.
3. Provide clear options for how to handle assessments that are done with some level of uncertainty. User generated comments linked to some level of uncertainty, could be color marked. To convey the uncertainty to the next user performing the evaluation.
4. Important to display the type of study used in the article exactly specified. To handle that type of study is sometimes difficult to assess. The type of study identified should then be displayed in the final report.
5. In addition to study type, the study's method and material used on the specific variant should be described in the comment field.
6. Write a short summary of the findings in the article related to the current variant.
7. Functionality provided for copying from the articles content and pasting into the comment field. The pasted text should be editable, since users sometimes need to shorten the pasted text.

8. There should be a spellcheck of how the authors are registered in the program when uploading articles. There is an issue in currently used systems that people mistype.
9. Handle future special cases where one reference is linked to and relevant for the classification of many different gene variants.
10. Color coding of evaluated references.
11. Providing the possibility to pause the classification process if needed and go back and retrieve text accidentally lost.
12. Like if one has finished work for the day and wants to continue the next morning. One of the most frustrating user experiences with the comment field would probably be if the user would lose the composed comment without any functionality for retrieving the text.
13. The list of articles should be sorted by some form of ranking criteria.
14. The user wants the articles marked as not relevant by the first person doing the evaluation to still be assessable for the second person doing the evaluation. Since there might be a disagreement upon the articles relevance between the first and second person doing the evaluation.
15. Important to display who that has done prior evaluations. Since the user will contact them if they need to discuss a prior assessment.
16. It should be registered in the evaluation if an article contains careless mistakes.
17. Insure that the all the articles presented in the list of references to evaluate mentions the variant.
18. Possibility to store other data than PDF references like conference posters if these were important for the classification.
19. Support easy access to articles that are recurring for many cancer genes that are addressed in the reference.
20. Accommodate a user preference for displaying issues with written information
21. Be critical towards how the structure of the comment field affects the comment formulated by the user. To insure that the user does not uncritically blend the article statements with subjective statements. To provide quality and make sure that there are no confusion when the next person is doing the evaluation.
22. Provide an option to select information not supplied, when information is missing in the articles.
23. To insure the validity of the various data entered into the evaluation, there should be specified requirements for the data. More specifically in regards of study type, method and material used.
24. There is identified a need for the system to handle demands brought on by differences in the variants being classified. Some type of variants is evaluated differently.
25. It should be displayed what type of study the reference is based on and if the evidence in the article is pathogenic/neutral or VUS. Sometimes it is the combination of references containing different studies that is decisive for the classification.

# 9 The paper prototype displaying the implications of the research on the system design

The research findings including the list of identified issues and themes are displayed in two paper prototypes. They are Photoshop redesign of a high fidelity prototype that was being developed. First is the list of references to be evaluated.

## 9.1 The first page of the prototype for article evaluation

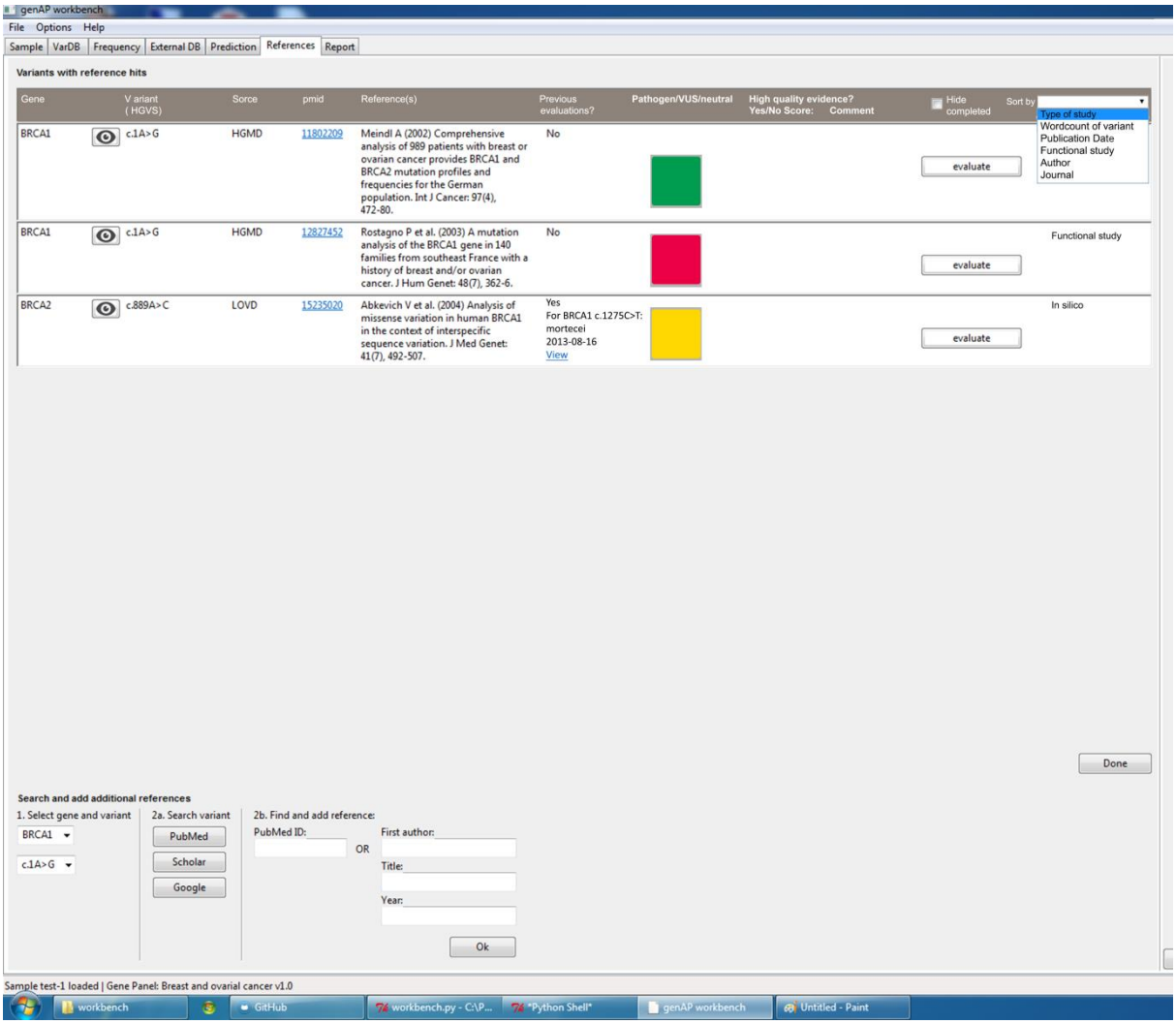



**Figure 26 the list of articles to be evaluated.**

The first page of the prototype for reference evaluation is presented in figure 26. The elements in the prototype are further explained in a presentation bellow.

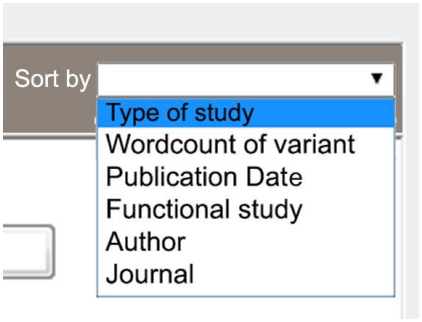



**Figure 27: Functionality for sorting of the list of articles to be evaluated.**

To support multiple user strategies to decide which article to evaluate first the list can be manually sorted by different selected criteria's. Like the type of study and word count (figure 27).


**Figure 28: To the left the type of study is displayed in the list. To the right is a color coding of evaluated articles.**

In figure 28 the evaluated articles are color coded with red meaning pathogenic variants, green neutral variants and yellow for VUS variants. This is based on the user developed method for evaluating articles and different user suggestions. In addition the top field is colored in a dark color with white text to highlight this area like a user suggested in the survey.

The type of study is displayed in the list at the far right side in figure 26. In the figure 28, this is show cased on the left side.

# 9.2 The second page of the prototype for article evaluation

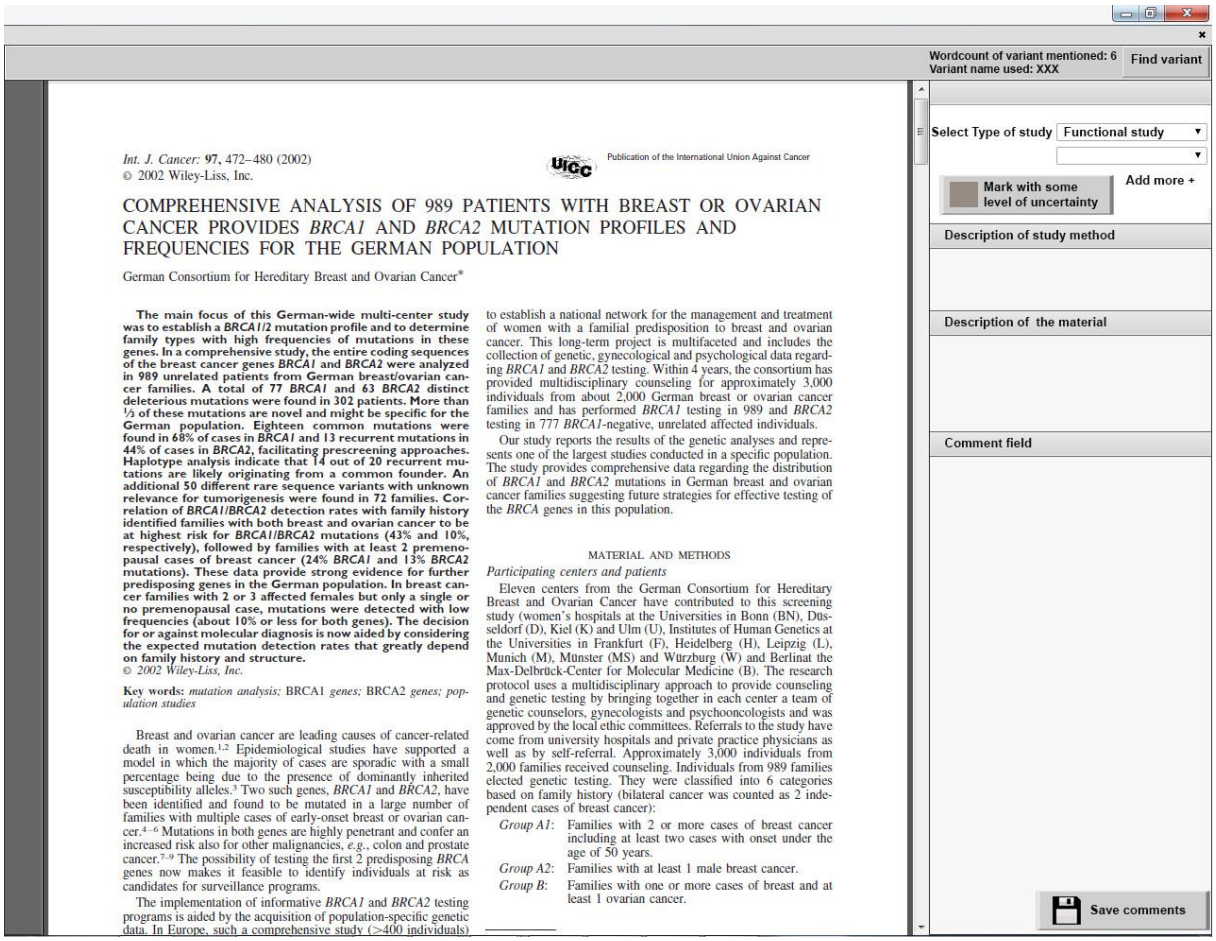



**Figure 29 The article evaluation functionality. (Meindl, 2002) used as an example article.**

When opening the PDF the comment field is provided, as shown in figure 29.On the top right side the variant word count is displayed together with the variant name used in the article. The button labeled "find variant" can be used to find the variant in the text.

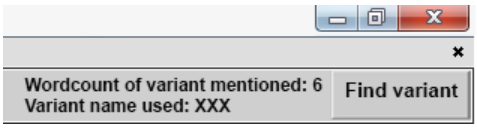



**Figure 30: The word count functionality.**

In figure 30, the "word count" functionality is shown.

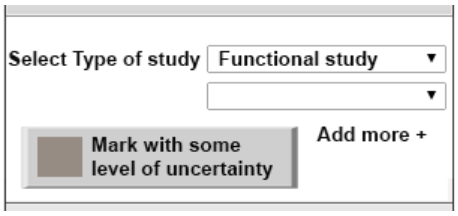



**Figure 31: The marking of text as statements with some level of uncertainties, and the selection of identified study type in the article.**

To handle issues of uncertainties, in figure 31 there is a possibility to click the button “mark with some level of uncertainty” then the selected text is colored gray, like in the user developed strategy to evaluate research references. The type of study can be selected in a drop down menu. The registered study type is later displayed in the list of articles to be evaluated. It is possible to register several studies for one article.

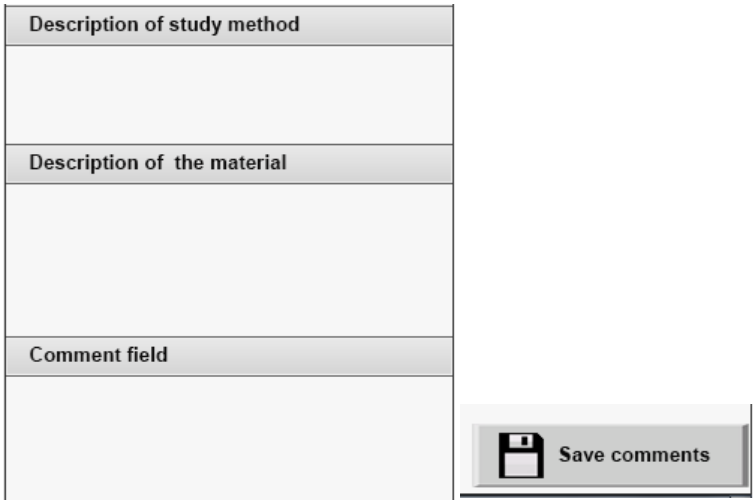



**Figure 32: The design of the comment field (to the left) and the save functionality (to the right).**

The comment field is divided into study method, study material and a comment field. This is shown in figure 32 on the left side.

It is added a button to save the comment since the user in the first user session stated that she wanted to have the possibility to pause the classification process if needed. This is presented in figure 32 on the right side. Like if one has finished work for the day and wants to continue the next morning. One of the most frustrating user experiences with the comment field would probably be if the user would lose the composed comment without any functionality for retrieving the text.

## 9.3 Evaluation

The next step would be to perform an evaluation of the paper prototype. This was unfortunately not possible. The workbench was being developed at the time of research and the article evaluation was not ready for user testing. The system will be user tested later and some of the findings from the user sessions will be explored further. The scope of the thesis

was to apply user-centered design and show how this approach can be used in such systems. On the same time the analyzed data from user sessions in this thesis were constantly informing the actual design of the system. The project has planned further user evaluations after the developer team has finalized the system design currently being implemented in Python. Many of the research findings in this thesis will be used both in the design of the system and addressed in later user testing's. These user tests will be past the time frame for this thesis research.

## 9.4 Summing up the results

By using user-centered design in the system development process, there were identified 24 design issues and themes for reference evaluation. In addition the nine most recurring issues in the data from the user sessions were identified in the analysis. This research also emphasizes the importance of deploying different user-centered design methods. How the user searches the PDF to see if and how many times the variant is mentioned in the article, was not mentioned by any user, it had to be observed. Then there was a need to purpose functionality to the user where the system automatically provides this "word count" at the time of choosing which article to evaluate first. The results from the survey clearly indicated that this was a desired functionality, which would probably not have been identified without observing the users performing the genetic classification. Another finding was related to the functionality that was highly valued by the user to find articles, where an automatically generated Google search was generated covering all the different names used for the variant in the literature. How this functionality was highly valued was mentioned in many of the user sessions. Through the observation it was observed how they had to search through the article many times to find the variant name used. This seemed like an ineffective solution, but it seemed since the user valued the functionality highly for other reasons this was not mentioned. By providing this automatically generated google searches and displaying the variant name used in the article. The system would add value to an already popular functionality.

# 10 Evaluation of the process

During the course of this thesis, I was focused on systematically applying the user-centered design approach and methods. They were an aid to maneuver in the complex domain of research and how to handle the massive amounts of issues uncovered in the user sessions. It was an ongoing challenge where I was constantly reflecting upon the design approach and my role as a designer. In the following chapter, I aim to present an evaluation of the design process in the form of a reflection upon the application of the user-centered design approach. In addition, I would like to discuss how this evaluation can contribute to future HCI research in complex domains mainly bioinformatics.

## 10.1 The user-centered design approach

There is a growing concern that HCI approaches are inadequate for complex, domain-specific systems (Chilana et al., 2010). In this thesis, the complexity is handled through a user-centered design approach. Two main challenges in the design process were how to maneuver in the complex domain of research and how to handle the massive amounts of issues uncovered in the user sessions.

The first challenge encountered was how to maneuver in the complex domain of research. This was handled through the user-centered design approach. The results from the user tests were the driving force behind the research and the key answer for how to maneuver in the complex domain. The next challenge was that the analysis of the data uncovered massive amounts of issues. Through the analysis, it was uncovered how some issues were briefly mentioned while other issues were recurring and repeating as a ”red thread” through the data. Issues that were briefly mentioned were also important, but often issues that were easily resolved. These issues were used in the development of the list of design issues and themes. When deciding which issues to address and validate in the survey. The focus was on “what are the users most focused on when provided the opportunity to influence the design in the user sessions?” This was answered by further exploring the identified recurring themes and issues identified in user session 1 (Figure 12). To investigate if, and how the issues were recurring in all the user sessions, the issues were sorted into themes and issues identified. These were presented in graphs (Figure 33) displaying if and how the issues were recurring in all the user sessions.

It was undoubtedly a considerable challenge to perform HCI research in the related complex research domain. The challenge was outweighed with the observation of how interesting and much needed HCI research is in the research domain. The designer’s lack of domain expertise is overcome by a focus on the user as a guide for the research and the design process. In addition, the lack of domain expertise is handled by teaming up with a domain expert with an interest of involving the user in the design process. The aim was to establish a room for negotiations in design between the designer and the domain expert; there were deployed two

initial user sessions with the already developed prototype together with the domain expert. This session was successful, and cooperation was established.

The relationship with the domain expert was a combination of Chilana et al. (2010), descriptions of iterative elicitation and persistent partnership in the initial user sessions. The partnership with the domain expert differed through the process of research. All analyzes of the research data were performed by the usability expert, and the results were presented and discussed with the domain expert. In the initial user sessions, the domain expert was involved in the planning of the sessions and co-facilitated the test sessions. After the two initial user sessions with the prototype, an observation was performed without the domain expert. The domain expert was present at the second observation and co-facilitated the following semi-structured interview. The analysis of the observations was performed by the usability expert. The interview with the lab doctor was co-facilitated with another researcher and the domain expert. After the observation, the researchers shared notes. A draft of the survey was written by the usability expert. This was further developed in collaboration with the domain expert. The research in this thesis is consistent with the findings of Chilana et al. (2010), showing that partnerships with domain experts leads to effective results. It was important for the researcher to acquire the necessary domain knowledge of genetics. The transcribed data from the user sessions contained much domain terminology, which was handled by performing Google searches in the analysis process. Thus, the analysis process increased the domain knowledge. It was also very beneficial to go back and reanalyze the data when the domain knowledge was higher. The lack of domain knowledge was also in some ways experienced as a benefit since the usability expert looked at the prototype with fresh eyes. More specifically the researcher's perspective on the prototype was gained from knowledge of the users. This perspective had fewer restrictions imposed by the domain.

## 10.1 The methodology

The focus of this research and the identified recurring design issues and themes were not on making generalizations, but studying how user studies can be applied in a complex domain. The interpretive research perspective has been important throughout the research in this thesis. The focus has been on understanding the design context through the eyes of the users. The interpretive paradigm is based on the assumption that reality can be viewed through "social constructions such as language, consciousness, shared meanings, documents, tools and other artifacts" (Klein & Myers, 1999, p. 69) In this research the document containing the user developed strategy for evaluating research references, was a good source for understanding how the users make meaning from the scientific research articles (It was presented in 8.3.8). This was used in combination with all the other research data collected, to discover and understand how users make meaning of different elements in the research article to contribute in the creation of a greater genetic knowledge. The article evaluation is in some ways highly subjective based on the users' previous experiences, individual preferences and their understanding of the domain. This is important to consider when designing for reference evaluation. Therefore, the interpretive research perspective was a good fit for this research.

## 10.2 The prototype as a research tool

As can be seen from Section 9, the role of, even the low fidelity paper prototype such as the one used in this thesis, was very important in making suggestions uncovered in Section 8 more concrete, and visible as part of the future system. In that way, the participants could argue further about the focus issue related to references. As the references were one of the most important issues for users, how references would be implemented as part of the system could still change their view about the problem. What became clear in this case is that flexibility in addressing the references is the real requirement the users had.

## 10.3 The access to users

Through the genAP project I was fortunate to have accesses to the future users of the system. Their working time is valuable, so it was important that the user sessions were well planned and executed within a planned time frame. The users were dedicated and positive towards contributing to the design of the system. In many of the sessions they also took time to explain domain issues on their own initiative, issues that were not easy to understand without an explanation.

## 10.4 Limitations

There are several limitations to this study. The proposed prototype is not evaluated, but is based on well-documented findings in the user sessions. The number of users is not statistical significant, but they are highly representative since all the users are part of a relative small group of future users.

# 11 Relating thesis research to relevant literature and theories

The research question in this thesis is "How can the use of user-centered design and user participation, considering EKP (emerging knowledge processes), design theory and actor-network theory framework as important in the complex context of genetic analysis platform design, support the design process of clinical genetic variant classification software?" In this chapter this is addressed by reflecting upon how ANT and EKP theory was beneficial in this research. In addition to discussing the potential other benefits not explored.

## 11.1 The ANT framework

In this thesis the ANT framework was used to analyze the development process of the clinical genetic variant classification software. This demonstrated how the use of ANT can support the design process by providing the researcher with a better understanding of the specific design process. ANT can be useful to get an understanding of the development of the system and the nature of the different elements that effect and are affected by the development of the system. With the equal inclusion of the human and non-human it gives a broad understanding of the design process. There were several important issues identified and examined. The application of ANT to the research case was limited to applying some selected ANT concepts. The deployment of a deeper and broader ANT-analysis could mediate a bigger understanding of the system development process. Through the use of the ANT framework the researcher is forced to be more critical and seek a deeper understanding of all the different elements. Just like the application of ANT to the research case in this thesis forced the researcher to be critical in the evaluation of all the elements related to the research. Reflecting on who and what that are affected by the research and identifying if some of these elements are ignored. This increases the awareness of issues that should be addressed in the design and development of the genetic classification software. It became apparent during the process of identifying the networks different components, that there were many elements the researcher had not taken into consideration. For instance how the user values supplementary online tables, but if the system does not support the uploading of other file formats than PDF, the system cannot store the supplementary online tables. The figure mapping out the network was growing over a couple of days as new elements were identified through the ANT process of reflections and applying ANT concepts to the development process. The focus of the researcher was forced to shift from a focus on the research findings and data analysis to a broader view examining all the different elements connected to or affected by the development of the system. The importance of including non-humans was especially apparent when considering how the article functions as a one-way communication medium that the user has to rely upon to understand the implications of the article's study on the variant classification. ANT can also be used to form a better understanding of the researchers' and the users' role in the development process.

## 11.2 EKP design theory

The findings of this thesis indicate a need for a flexible system that supports diverse user strategies. Some of the user strategies are deployed to handle the emergent knowledgebase of genetic research. Like when more genes are found to be the cause of a disease, than the genes already predefined in a specific gene panel. To provide flexibility is also important from an EKP theory view, where flexibility is needed to support the contextualization that aids in the comprehension of emerging knowledge processes. The genetic classification process requires tacit and explicit knowledge, general and contextual knowledge. Knowledge and expertise is also required in applying the knowledge. Therefore the genetic classification process can be seen as an emergent knowledge process. This leads to the argument, that the 6 principles for the design process of EKP support systems can be valuable for the design process in this thesis. The first principle is to design for customer engagement by seeking out naive users. All the participants in this research have high levels of expert knowledge, but there is a distinction between the lab doctors and the lab engineers. The lab doctors possess more knowledge of advanced genetics. Most of the participants in the user tests are lab engineers. To further use this principle more naive users could be involved in user sessions. EKP design theory argues that it is almost impossible to predict in advance who will be the future users of the system. So the system should not target specific user roles. The second principle is to design for knowledge translation through radical iteration with functional prototypes. This principle was actively used in the development of genAP workbench and proved to be beneficial. The third principle is to design for offline action. The issue most relevant for this is the verbal discussions with local colleagues performed regarding uncertainties in the classification process. This is described in this thesis. Further support for offline action could be important especially if the system is put in use in other hospitals. The fifth principle is to design for implicit guidance through a dialectical process. The sixth principle is that the system must be extremely flexible; developers should componentize everything, including the knowledge-base. The overall system design is split into separate components and supporting flexibility by easy traversing between the different tabs with the components. The knowledge base of the variant classification results are computerized and stored in a database.

### 11.2.1 Summing up

Both ANT and EKP displays how important it is to develop systems that includes a more broad user group and variant case diversity. The robustness of the system will be defined in how well it handles anything that does not fit in a predefined notion of the average user and the average variant classification case.

Focusing on designing systems that looks beyond the design for the average case or average user. Since from an EKP theory view these are changing and the development should not be restricted to specific user roles or restricted to specific genetic knowledge that is emerging

and changing. ANT showed how important it is that all the different variant cases and different users are included in the system design.

Based on how well the EKP design theory fits the complex context of genetic analysis platform design, the theory is beneficial to consider in the design process of such systems. The application of ANT on the complex context of genetic analysis platform design facilitated a deeper understanding of the design context. This indicates how both ANT and EKP design theory can support the design process of clinical genetic variant classification software.

## 11.2 Relating thesis research to relevant literature

The research in this thesis highlights the importance of providing the users with support for multiple user strategies and drawing upon expert knowledge through user involvement. Research on clinical decision support systems indicates that a crucial factor for its success is to present the right information to the user at the time of decision making. But how is this achievable? In changing landscapes like bioinformatics where the bases for the decisions are changing and evolving as scientific research uncovers new knowledge. The research in this thesis further support the findings from others indicating that evidenced-based use of genetic data demands a need for information technology to support the management of the rapidly changing and growing genetic knowledge base (Hamburg & Collins, 2010; Neri et al., 2012; Scheuner et al., 2009). This leads to an argument for the value of drawing on established theories and methods from the field of knowledge management. Alavi and Leidner argue that due to "knowledge management's complexity, the appropriate approach(s) of knowledge management, resource requirements and software tools are reflected in the characteristics of the specific knowledge management process" (Alavi & Leidner, 2001, p. 131). This is also relevant for the findings in this thesis since the knowledge management demands brought on by the introduction of evidence-based use of HTS sequence data, is highly unique, research dependent and will demand new knowledge management approaches, resource requirements and software tools. To meet these demands the knowledge management needs to be shaped by the characteristics of the diverse user knowledge (both tacit and explicit) and strategies deployed in the bioinformatics knowledge management process. The research in this thesis further argues that to develop systems that meet these demands, the users working with the variant classifications needs to be valued as the knowers and only source for this information. For the design of such systems the users' needs to be involved in the design process, as argued for by others (Bolchini et al., 2009; Chilana et al., 2010; Javahery, Seffah, & Radhakrishnan, 2004; Neri et al., 2012; Pavelin et al., 2012; Shyr et al., 2014).

The research in this thesis also argues for the importance of visualizing the actual content of the articles in the list of articles, before the user opens them. This further support the findings from Bolchini et al. (2009). The design requirement that the list of articles should be sorted by some form of ranking criteria, which also could be manually selected. Is also argued by

Bolchini stating that “it is important to explicitly communicate to the user the actual ranking criteria used for displaying the results (as results are displayed) and, possibly, to allow sorting the obtained results by multiple, additional attributes (e.g. by publication/release date, by alphabetical order)” (Bolchini et al., 2009, p. 412). The ranking criteria is further explored in this thesis to consist of “type of study”, “word count of variant name”, “publication date”, “functional study”, “author”, and “journal”. Shyr et al. (2014) also argue for the importance of giving the user options to customize the type and order of information presented. This is consistent with the findings in this research.

Through the user sessions with the prototype it was evident that the users valued that the system should support the users workflow. This is in consistency with Shyr et al. (2014) findings. The user also valued that the system incorporated all external online resources used in the process of annotating the variants. This is also consistent with Shyr et al. (2014) findings. They state that the task of accessing all the external online resources in the process of annotating the variant was reported to be amongst the most time-consuming steps in the interpretation process (Shyr et al., 2014).

The user session 1 revealed several issues related to the interface design elements of the prototype system, which was then fixed. These issues concerned among others button labeling, language, placement and organization. In the following user session, no such issues were found indicating that these usability issues were identified and easily fixed. This is consistent with the findings of Neri et al. with their argument that such “results demonstrate that using a development and design process that is user focused helped optimize the value of this application for personalized medicine” (Neri et al., 2012, p. 950).

The importance of user involvement in the development of clinical decision support tools is also evident in how local work practice often is unique. Lindgren argues that “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 use environment is assessed, and requirements for a CDSS are formulated”(Lindgren, 2011, p. 129). The many characteristics of the local work practice are also shown in this research, how the characteristics benefit from system support is focused on and in some ways addressed in the list of design issues and themes. The local collaboration that consists of verbal discussions is not currently supported by the system. Many of the users did not see any benefit from supporting online discussions since there were locally available colleges. One user was focused on limiting the need for verbal discussions, by marking uncertainties in the text. The solution of highlighting in gray statements that has some level of uncertainties, warrants further system support for collaboration. The prototype system also supports collaborative analysis, developed based on user participation. The process of annotating the gene variants in the system is done by minimum three users consecutive. They comment on individual articles and all the information is available to the users involve in the classification process. The support of collaborative analysis is positively received by the users in the user sessions with the prototype. This is consistent with Shyr et al. findings 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)

The users handle the different gene variant classification cases by deploying diverse user strategies. These strategies should be addressed in the design of the system to accommodate the need for presenting the right information to the user at the time of decision making. Some of the strategies are built upon individual preferences, while others are brought on by the different demands from variant classification cases. As the knowledge bases are changing most likely the user strategies will alter or increase especially to handle special variant classification cases. Some variant classification cases that had other demands than the average variant classification case was identified in this research. One of the users address this by stating the "system has to be flexible." The research in this thesis indicates that there is no uniform user strategy used in the task of evaluating research articles. Shyr et al. warn "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). The need for flexibility and handling unique cases is also found in the data from this research. To achieve a system design that supports the user in an efficient way the research in this thesis argue for basing the system on user knowledge of the multiplicity in how the user decision process is performed. This user knowledge should be gathered through user involvement in the design process and with HCI design approaches and methods.

# 12 The road ahead

The application of ANT to the research case was limited to applying some selected ANT concepts, the deployment of a deeper and broader ANT-analysis could mediate a bigger understanding of the development of the system and the nature of the different elements that effects and are effected by the development of the system

But also looking beyond user-centered design and applying other HCI approaches like participatory design (PD) and service design. PD could be beneficial amongst others due to the aspect of mutual learning. A usability expert needs to acquire knowledge of the domain and the users' needs knowledge about the possible technological solutions to develop useful and new design solutions. The users performing the invaluable work and improving patients' health deserve to have a say in in the development of systems that will affect their work environment.

Most of all users should be provided with systems that support their work better. To achieve better support, the users need to be involved in the design process. The successful application of user-centered design in this research hopefully inspires others to follow and adopt a HCI design approach in the future development of bioinformatics systems.

# 13 Conclusion

This research gives a unique contribution to the field of HCI research, by displaying how user-centered design is both applicable and benefited to system development in the complex domain of bioinformatics. Through the application of the ANT framework the role of the user in the complex domain was identified as a crucial source of information for the design process. The role of the user identified in the ANT analysis illustrates how the USD perspective can be beneficial in bioinformatics. The application of user-centered methods revealed how the strength of these methods results in valuable design concepts that cannot be gathered simply through conversations with the users or by studying workflow charts. The introduction of USD and HCI approaches in bioinformatics is an important step in the process of dealing with the "Big data" and analyst bottleneck problem in bioinformatics. The findings in this thesis indicate that the combination of the approaches used may be a good way of overcoming usability challenges when working in complex domains. User-centered design in combination with actor-network theory, design theory and emerging knowledge processes get a deeper understanding of the situated design context. This research suggests a needed shift in how the user and user involvement is regarded in the design process of genetic variant classification systems and other bioinformatics systems.

# References


Aanestad, M. (2003). The camera as an actor design-in-use of telemedicine infrastructure in surgery. *Computer Supported Cooperative Work (CSCW), 12*(1), 1-20.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107-136.

Baligh, H. H., Burton, R. M., & Obel, B. (1996). Organizational consultant: Creating a useable theory for organizational design. *Management science, 42*(12), 1648-1662.

Bolchini, D., Finkelstein, A., Perrone, V., & Nagl, S. (2009). Better bioinformatics through usability analysis. *Bioinformatics, 25*(3), 406-412.

Boyd, S. D. (2013). Diagnostic applications of high-throughput DNA sequencing. *Annual Review of Pathology: Mechanisms of Disease, 8*, 381-410.

Chilana, P. K., Wobbrock, J. O., & Ko, A. J. (2010). *Understanding usability practices in complex domains.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., & Walters, L. (1998). New goals for the US human genome project: 1998-2003. *Science, 282*(5389), 682-689.

Cressman, D. (2009). A brief overview of actor-network theory: Punctualization, heterogeneous engineering & translation.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*: Harvard Business Press.

Davies, J., Harmelen, F. v., & Fensel, D. (2002). *Towards the semantic web: ontology-driven knowledge management*: John Wiley & Sons, Inc.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (Vol. 2): Wiley New York.

El Sawy, O. A., & Bowles, G. (1997). Redesigning the customer support process for the electronic economy: insights from storage dimensions. *MIS quarterly*, 457-483.

Finken, S. (2013a). METHODS 1: OBSERVATION, INTERVIEW, DOCUMENTS. 2014, from http://www.uio.no/studier/emner/matnat/ifi/INF5220/h13/lecture3-2.pdf

Finken, S. (2013b). Overview of the field.

Gallon, M. (1999). Some elements of a sociology of translation. *The science studies reader*, 67.

Gherardi, S., & Nicolini, D. (2005). Actor-networks: ecology and entrepreneurs. *Actor-network theory and organizing*, 285-306.
Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM, 28*(3), 300-311.
Grint, K., & Woolgar, S. (2013). *The machine at work: Technology, work and organization*: John Wiley & Sons.
Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine, 363*(4), 301-304.
Hannemyr, G. (Autumn 2013). ( Internet) Research: Processing personal data and the law

(And maybe something about citation). from http://www.uio.no/studier/emner/matnat/ifi/INF5220/h13/research11.pdf
Hutchin, E. (1991). The Social Organization of Distributed Cognition. I: Resnick: LB.
Javahery, H., Seffah, A., & Radhakrishnan, T. (2004). Beyond power: making bioinformatics tools user-centered. *Communications of the ACM, 47*(11), 58-63.
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher, 33*(7), 14-26.
Kakabadse, N. K., Kakabadse, A., & Kouzmin, A. (2003). Reviewing the knowledge management literature: towards a taxonomy. *Journal of knowledge management, 7*(4), 75-91.
Kassahn, K. S., Scott, H. S., & Caramins, M. C. (2014). Integrating Massively Parallel Sequencing into Diagnostic Workflows and Managing the Annotation and Clinical Interpretation Challenge. *Human mutation, 35*(4), 413-423.
Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj, 330*(7494), 765.
Khoury, M. J., Gwinn, M., Yoon, P. W., Dowling, N., Moore, C. A., & Bradley, L. (2007). The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genetics in Medicine, 9*(10), 665-674.
Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS quarterly*, 67-93.
Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*: Harvard university press.

Latour, B. (1990). Technology is society made durable. *The Sociological Review, 38*(S1), 103-131.
Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems practice, 5*(4), 379-393.
Law, J. (2009). Actor network theory and material semiotics. *The new Blackwell companion to social theory*, 141-158.
Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research methods in human-computer interaction*: John Wiley & Sons.
Lim, Y.-K., Stolterman, E., & Tenenberg, J. (2008). The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI), 15*(2), 7.
Lindgren, H. (2011). Towards personalized decision support in the dementia domain based on clinical practice guidelines. *User modeling and user-adapted interaction, 21*(4-5), 377-406.
Madden, R. (2010). *Being Ethnographic: A Guide to the Theory and Practice of Ethnography*: SAGE.
Manolio, T. A., Chisholm, R. L., Ozenberger, B., Roden, D. M., Williams, M. S., Wilson, R., . . . Eng, C. (2013). Implementing genomic medicine in the clinic: the future is here. *Genetics in Medicine, 15*(4), 258-267.
Mardis, E. R. (2010). The $1,000 genome, the $100,000 analysis. *Genome Med, 2*(11), 84.
Markus, M. L., Majchrzak, A., & Gasser, L. (2002). A design theory for systems that support emergent knowledge processes. *MIS quarterly*, 179-212.
Meindl, A. (2002). Comprehensive analysis of 989 patients with breast or ovarian cancer provides BRCA1 and BRCA2 mutation profiles and frequencies for the German population. *International journal of cancer. Journal international du cancer, 97*(4), 472-480.
Murdoch, J. (1998). The spaces of actor-network theory. *Geoforum, 29*(4), 357-374.
Neri, P. M., Pollard, S. E., Volk, L. A., Newmark, L. P., Varugheese, M., Baxter, S., . . . Bates, D. W. (2012). Usability of a novel clinician interface for genetic results. *Journal of biomedical informatics, 45*(5), 950-957.
Neuendorf, K. A. (2002). *The content analysis guidebook*: Sage.
Pavelin, K., Cham, J. A., de Matos, P., Brooksbank, C., Cameron, G., & Steinbeck, C. (2012). Bioinformatics meets user-centred design: a perspective. *PLoS computational biology, 8*(7), e1002554.
Plon, S. E., Eccles, D. M., Easton, D., Foulkes, W. D., Genuardi, M., Greenblatt, M. S., . . . Tavtigian, S. V. (2008). Sequence variant classification and reporting: recommendations for improving the interpretation of

cancer susceptibility genetic test results. *Human mutation, 29*(11), 1282-1291.
Preece, J. R., & Rogers, Y. (2007). SHARP (2002): Interaction Design: Beyond Human-Computer Interaction. *Crawfordsville: John Wiley and Sons, Inc. Answers. com Technology*.
Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: Beyond Human - Computer Interaction*: Wiley.
Sanders, E. B.-N. (2002). From user-centered to participatory design approaches. *Design and the social sciences: Making connections*, 1-8.
Sanders, E. B.-N., & Chan, P. K. (2006). Emerging Trends in Design Research.
Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-design, 4*(1), 5-18.
Scheuner, M. T., de Vries, H., Kim, B., Meili, R. C., Olmstead, S. H., & Teleki, S. (2009). Are electronic health records ready for genomic medicine? *Genetics in Medicine, 11*(7), 510-517.
Shyr, C., Kushniruk, A., & Wasserman, W. W. (2014). Usability study of clinical exome analysis software: Top lessons learned and recommendations. *Journal of biomedical informatics*.
Silverman, D. (1998). Qualitative research: meanings or practices? *Information systems journal, 8*(1), 3-20.
Silverman, D. (2005). Doing Qualitative Research. *Sage Publications*.
Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang, P. C. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association, 8*(6), 527-534.
Stanley, C. M., Sunyaev, S. R., Greenblatt, M. S., & Oetting, W. S. (2014). Clinically Relevant Variants–Identifying, Collecting, Interpreting, and Disseminating: The 2013 Annual Scientific Meeting of the Human Genome Variation Society. *Human mutation, 35*(4), 505-510.
Tatnall, A., & Gilding, A. (2005). Actor-Network Theory in Information Systems Research.
Torkamani, A., Scott-Van Zeeland, A. A., Topol, E. J., & Schork, N. J. (2011). Annotating individual human genomes. *Genomics, 98*(4), 233-241.
Van der Velden, M., & Culen, A. L. (2013). Information Visibility in Public Transportation Smart Card Ticket Systems. *International Journal On Advances in Networks and Services, 6*(3 and 4), 188-197.
Welch, B. M., Loya, S. R., Eilbeck, K., & Kawamoto, K. (2014). A Proposed Clinical Decision Support Architecture Capable of Supporting Whole Genome Sequence Information. *Journal of Personalized Medicine, 4*(2), 176-199.
Wiberg, M., & Stolterman, E. (2014). *What makes a prototype novel?: a knowledge contribution concern for interaction design research.* Paper

presented at the Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational.

# Attachment

A: The Report from the first two user sessions

B: The consent form sent out with the survey

C: The survey

D: Pictures from the content analysis

E: Pictures from the data analysis

# A: The Report from the first two user sessions

# Report from the two first user sessions with GenAp workbench

## User session 1

### Changes in work tasks

The prototype is based on the users current SOP (standard operation procedure), but also some changes are made for the work-tasks to meet requirements brought on by the shift from Sanger sequencing to HTS sequencing. The biggest challenge from introducing HTS technology is the massive amount of sequence data. The domain expert states that there is «a focus on the scalability of the system, because suddenly the amount of data will explode and you will need to have a system that is robust enough to handle 14.000.000 gene variants."(User session 1) When the user interacts with the prototype the analysis uncover that some of the reactions from the user is based on the changes made to their work tasks. The user is reflecting on how this differs from their current work practice. Then the user is engaged in discussions and negotiations with the domain expert on how these solutions should be designed to be optimal for their work practice.

### Sanger sequencing issues

The user states that when «using the program for Sanger sequencing data there can be an issue with delays, since BRCA takes longer time than other genes since we have so many tests for BRCA. There will be a time delay in relation to what comes out of the finished results» This delay is not an issue with HTS since all the data is received simultaneously.

### Negotiations

The analysis of the data from the user session revealed that there was a great deal of negotiation taking place. As mentioned some of these negotiations were brought on by the demands from the new technology, but the negotiations were also between the stakeholders' interests and user preferences. Most of the negotiations resulted in compromises that were agreed upon. One of the initial issues was related to a functionality that was not provided in the current prototype of the system. More specifically the quality control (QC) of the sequence data generated from the Next generation sequencing (NGS) technology. Basically

she wanted a user-friendly NGS QC Tool for quality check and filtering to get high-quality data for the analysis. The user explains that this process is currently done manually on paper. The raw sequence data results are printed out and the lab-clinician looks on every variant on the list and decides which variants that is relevant. The user states "What we currently do is that we printout, then we inspect every gene variant. And say no, this was nothing. No, this was nothing. Yes, this is actually a heterozygous. No, this was nothing, a homozygote, but not already classified as normal" (Participant statement). It is described as a labor-intensive task, where they typically begin with a list that's 1 1/2-2 pages long and after filtering out the data that is not relevant (like the technical artifacts) they are left with five variants. Of these five some are normal variants. The length of the first list of variants prior to the QC is highly fluctuating dependent on the quality of the sequencing. The user explains that "With a good sequencing. Then maybe 10 gene variants per test sample and for a slightly worse, maybe it could be 40 gene variants." The user states that "I would have liked to have some kind of sequence program, where you get all of the gene variants and are able to select acceptance or discard through the entire list of variants."

Through the discussion with the domain expert it is stated that this functionality is not currently in the scope of the prototype. The QC of the data and the construction of the list that is uploaded into the workbench are supposed to be done prior to using the workbench prototype. The user states that for them to just select which variants they want to test is much more efficient. The time used to first create the high quality data gene variants list and then uploading it into the workbench followed by a process of filtering out the normal variants pre classified in-house is much more time consuming. The user states that the way she currently work with the Sanger program «it would have been much faster just to type in I want to test this, this and this variants. And I know that those other gene variants are normal variants. » When the user works with gene variant testing for breast and ovarian cancer, this could be done by selecting only the variants related to the genes BRCA1 and BRCA2. Then the user states that there would "For then maybe it are just two, three gene variants." Currently when the user finds a variant then they go to the in-house database and check if it has a prior valid classification and if it is a 5 then they don't have to analyze it. But in the prototype the user have to quality check and make the list and upload it to the program before seeing that it is a valid 5 classification. In the discussion of this functionality it is stated by the domain expert that one of the reasons for augmenting the work-task of uploading the gene variants into the

system is because the stakeholders need to upload the entire list of variants into the in-house database (VarDB).

### Desire to test multiple samples simultaneously and a Need for flexibility

One of the clearly stated demands from the user was that it should be easy to get the gene variants into the workbench. Another related issue that was discussed thoroughly was the use of sample-id or project-id when uploading the high quality sequence data into the workbench. The prototype was developed to use sample-id for the uploading of sequence data. The user addressed a preference for using project-id, since then they could upload many test- samples at once. This is reflected when the user is asked what kind of program that had been ideal for her? She replies that "when they set up data for instance for BRCA in the Sanger program they set up 12 tests per project. BRCA1 and BRCA2 are set up together. 12 separate patients and the same patients for BRCA1 and BRCA2 in parallel. Then you go through the project". She states further "If I could take all of those together and select this patient has this and this variant, this patient has this and this variant, test them all at once. That would have been really good." She elaborates that to "not having to sit and select sample-id all the time for 12 sample-tests. Instead just select test all the 12 at once. "She states further "I would like to make a list by selecting which sample and which gene variant I am interested in and then upload this list." To be able to do this efficiently she wants to fetch the variants for the analysis by using project-id instead of sample-id and then test all the variants in the project simultaneously. The Project-id is a unique name. It is created by combining type of analysis + date + who has run the project (initials) + a digit. The combination of all of this various elements results in a Project-id that is very long.

The user states that there are more problems with only using sample-id. The user explains how there is a need for flexibility in the choice of how you select to begin the analysis. The user is currently doing the analysis on a selected gene basis she state "The way it is done today is that we take it continuously as it comes out of the Sanger sequencing. One and one gene variant at a time." The user explains that "we have divided the analysis into the BRCA genes and Lynch Genes. BRCA goes as a set 1 and 2. The Lynch gens goes as a set with three. And then goes all the other genes mixed. "She further explains that "If I were to work with other genes. Then I would collect an amount of samples that I get to fit on 10 times the plates. And then I label it as a project- x. Then I run it as a project and analyze the project. "She states that project-id could be used, but they have long names.

When HTS is used in the prototype the intent is to have all the genes at once. The domain expert states that “It might provide some benefits to analyze multiple genes at once. You might expect to just find one mutation. So the analysis of all genes together, in a way provides some advantages in weeding out the suspected pathogen.” In the prototype when the sample-id is used to upload the tests they are linked to a specific gene panel. These gene panels can be specified for both Sanger sequencing and HTS sequencing data. Currently the users are using gen panels for BRCA genes. In the workbench there is going to be made more specified gene panels that can also just consist of one gene. The user states that still “somewhere there should be an possibility to select which gene I'm interested in ” She furtherer states that” it has to be flexible” to handle special cases, like if there are combinations in the *genetic testing.* She elaborates that “frequently there are combinations where the requisition is for BRCA and then later the requisition is for another gene (like PTEN) using the same sample-id.” So the same sample can be used for many gene panels. There are also cases where the requisition is only for BRCA1, because BRCA2 has been done in a prior analysis. The user states that to meet these different demands « the system has to be flexible” Through the negotiation and discussion of this topic the domain expert and user agree that there should be provided an option for selecting gene, when the program is used for analyzing Sanger sequencing data.

### Other factors relevant for classification

The user states that” we had a disease-causing gene variant in gene codon 335 a stop codon and it was classified as disease-causing since it was well documented. And then we had another variant in stop codon 336, the next codon and it was not well documented, so it was only classified as a class 4 gene variant. But you can imagine that if you have one in 335 that is disease-causing, why should not it be that in 336? » She further states that If we just get that good article that states that «here the gene domains cannot withstand anything. » She wants it registered in the system. The domain expert agrees with the user that this should be accounted for. The only question is when the research is mature enough. The user states that “it is said that BRCA1 cannot withstand anything” but BRCA2 is the exception that can handle mutation in the stop codon in the last exon. This illustrates how the system has to be flexible and handle both emergence in knowledge of genetic classifications and special cases that are exceptions.

### Traceability

Another related issue that the user addresses is the problem of traceability. There is a need for information about which date it was run, which batch and by whom. This should be displayed in case one has to troubleshoot backwards. Another problem for the traceability is that sometimes tests are rerun, these run in another project with a new project-id. An advantage by doing so is that you are sure that you have not made an error of switching tests. When two people pick out the test-sample independent of each other chances for doing an error are small. Tests are rerun if they find a disease-causing gene variant to confirm. Other reasons for reruns are bad quality sequence data or to confirm the classification of a VUS.

The user states «I think that if you ever would like to go and search backwards or decide to do scientific research. The user states "Say I have a VUS and I decide I want to actually start and work on a research project on this VUS. If I want to then find the raw data, then I must know where it is."(User session 1)Today such information is kept in the Swiss lab program. There it is recorded which project the test-sample is run in. If you just check the VUS it is not certain that you will find all the test-samples that relate to the VUS.

## Emergent demands

The user also asks the domain expert "what is possible to retrieve from the in-house gene variant classification database?" She argue that it is important to think about in advance how easy it will be to retrieve data from the database and states "It will be a demand for extracting something from the database and that is not always so easy to do. When you have made the database everything is fine, but then comes the questions. Ok, but maybe you could get out this and make some calculations and so on. Just be prepared for that."

## The current flow of the system

The user is positive towards the use of tabs and the possibility to navigate between tabs. This makes the process more flexible and it is possible to traverse the tabs as needed through the analysis. The prototype supports the analysis by providing a functionality where some gene variant can be accepted as a classification and is then not displayed on the following tabs but are put in the last tab where the report is created. The variants that are not accepted are displayed in the following tabs. This got positive feedback from the user especially the possibility of traversing backward to specific tabs.

### Visual design requirements

Some of the results from the user session gave insight into concrete design requirements. Visually the user wanted to increase the font size used throughout the system. Especially the most important information should be bigger. The use of color coding in the program got good feedback. Only a slight remark that the user personally did not like the pink shade used. She stated that she would like to add more color to the overall interface design. The buttons and options should also be bigger.

### Functionality requirements

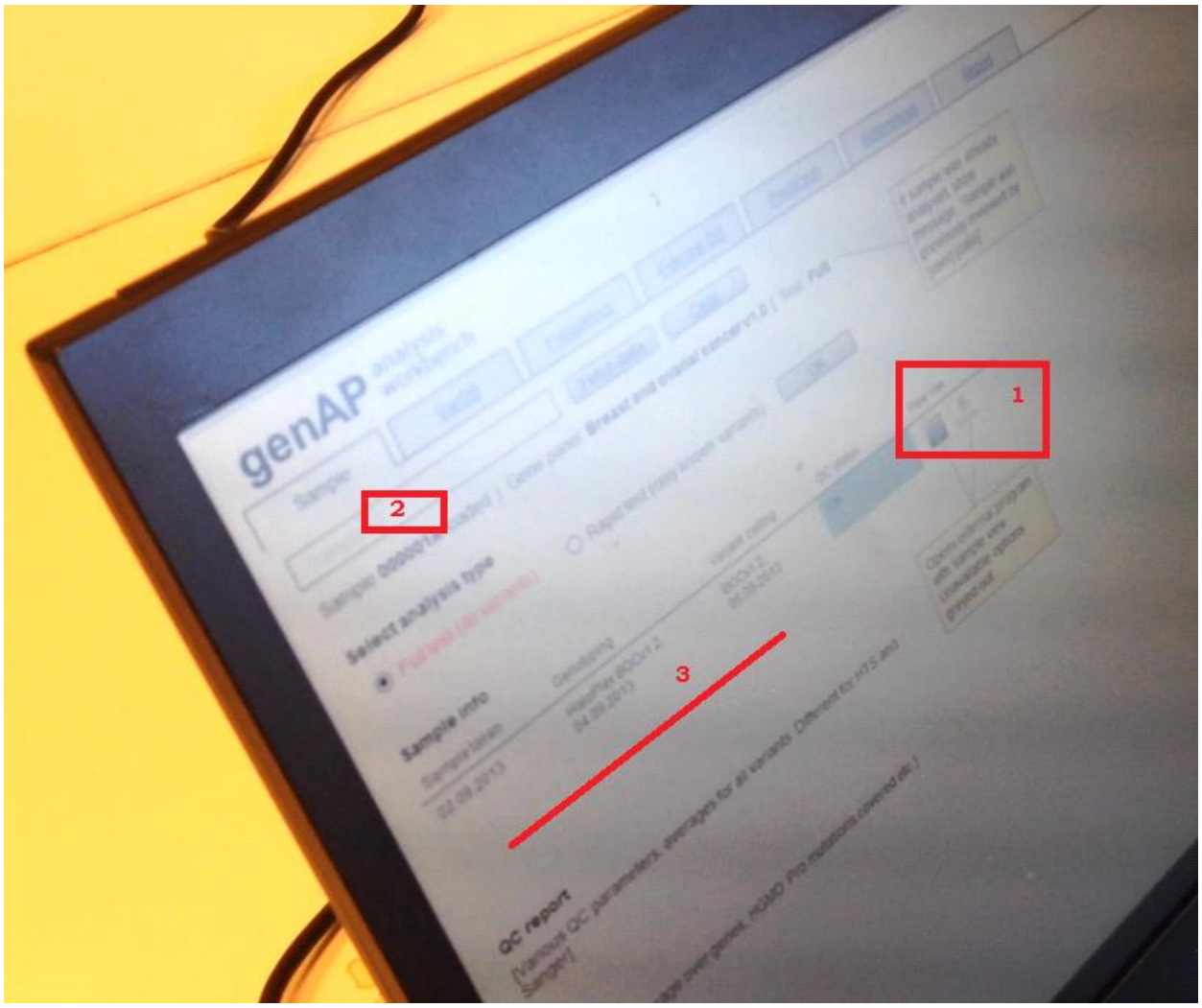
It is desirable to provide functionality to take a break in the process and then resume later. Some sort of tracing the annotation process. The user asked "what do you do if you are uncertain at some point of the analysis? Do you just stop? How the program deals with this issue is important." The user wants to be provided with a clearly defined policy for how to handle such a situation.

### Ease of use

Some of the feedback was focused on *increasing* the perceived *ease of use* of the system. The user stated that the whole process of annotating the gene variants in the system should be distinct and easy. To get the list of gene variants uploaded to the system should be an easy task. The user states that it is important that the end result from the system is something good.

### Results from video analysis

## Page 1: Findings related to the sample information.




Picture 1: Sample information tab, the numbers in the picture is described in the list bellow.

1. The user would like a QC- tool as part of the program. But stated that if such functionality was not provided and the list is made elsewhere prior to uploading into the workbench, then there are no need for a functionality in the workbench to look at the raw data, since the user have already looked at that in the Seqscape program prior to starting the analysis in the workbench (Marked as 1 in picture 1).
2. In the search field the user wanted to add the option of selecting gene in addition to sample-id (Marked as 2 in picture 1).
3. Sample information: Here the user wanted to be provided with more information. Especially information about which project the sample is related to if using sequence data from Sanger. Since that information is needed to be able to troubleshoot backwards and offer traceability. The information should be witch batch and which day. This information should be in the metadata, but the user states that she wants it to be provided here too (Marked as 3 in picture 1).

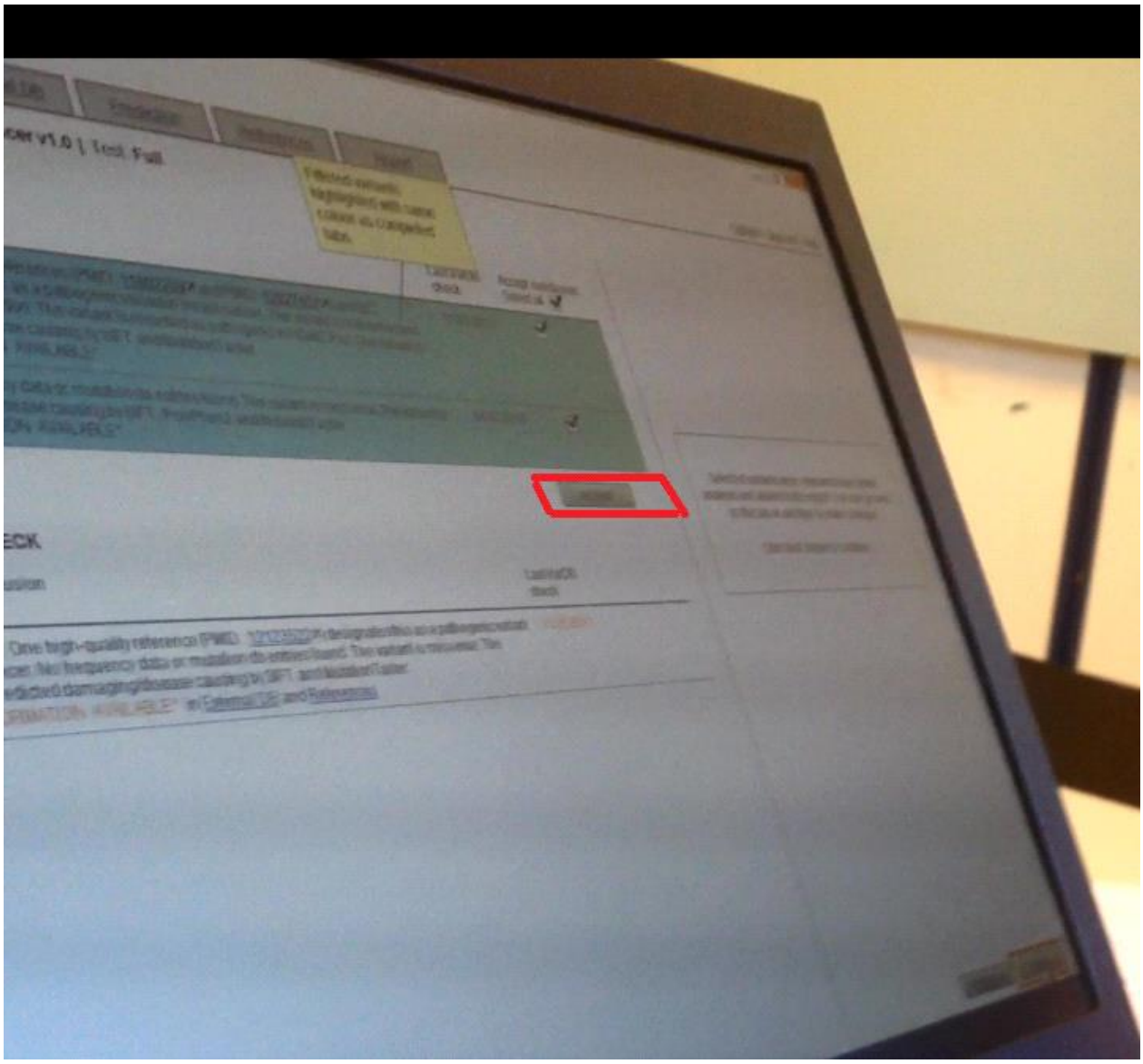
Overall and especially for this first page the user request a clear visual representation of what has been done in the analysis of this data prior to the users current analysis? More specified display what has been done so far (by others) and where you are currently in the process.

## Page 2: Findings related to the in-house database VarDB

Now the user has to manually look up if the variant is already classified in-house. In the prototype information from in-house classifications is integrated into the program through the database VarDB. The user is positive towards this functionality.

From the video analysis it is indicated that the buttons in the bottom right, should be visually displayed better.

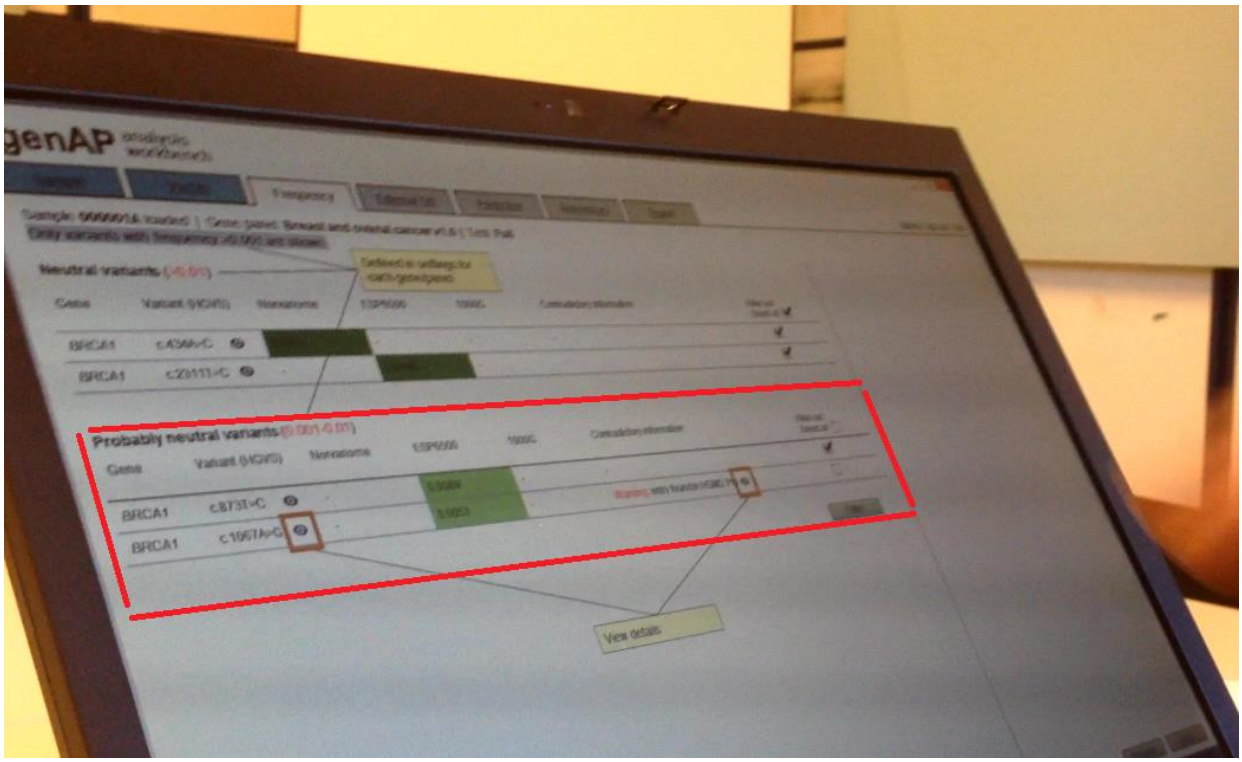
It is further stated by the user that the class 1 variants does not need to be displayed in this page. The user is positive towards that the system shows if there are new information available.



Picture 2: The information tab related to the in-house database.

## Need for constraints

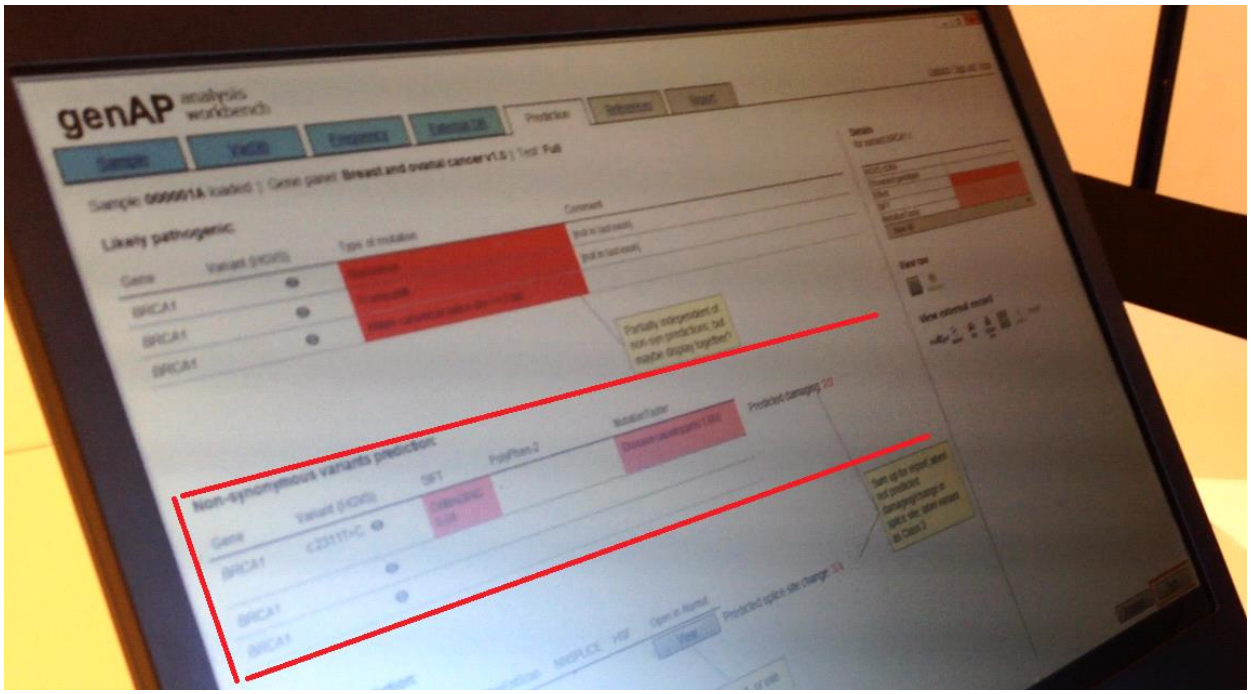
The user was confused by the buttons in the middle and on the bottom right. It was not evident which that should be used at which time in the process. The user states that the button in the middle should be grayed out and not assessable when appropriate.




Picture 3: The frequency information tab.

The user states «that I am really glad you use frequencies and not percentages here. » There is a need for providing more information on this page, since the user wants to add information about genotype frequency.

The button labeled «filter» was perceived as confusing when the other button from previous tab was labeled «accept». There is a need for consistency in the design of the program. The functionality should be indicated more clearly in the choice of labels on the buttons. Since punching on the button results in the variants not being displayed in the following tabs, but is put in the last tab (report). The user suggested labeling the buttons with “take out of analysis» or something that visually shows the functionality. Another button that the user perceived as confusing was the «previous» button she did not think it was evident enough by this labeling that it was just going backwards and not undo.




Picture 4: The information tab related to prediction results.

The user states that the red marked area in the picture 4 is not something the clinicians uses in the analysis. It was earlier included in their standard operation procedure (SOP), but not in the current SOP because this is uncertain predictions. Based on a negotiation with the domain expert it is agreed upon to keep this functionality but only visually displayed for the variants this is relevant for. This could be done by hiding it for certain panels.

The prototype displays splicing predictions (in the bottom field) which the user found to be valuable. There is also a need for providing more information on this page. It should show information about protein domains, how conserved it is.

## The report page

For the report tab in the prototype the design was not finished at the time of the user session. The user got an explanation of the intended content. «There will be much information at once. The most important things are placed at the top, with color codes. Also you get a summary of what you have done in these previous tabs. If the findings of two references of high quality, both are pointing in the direction of pathogenic, then it will be proposed to classify it as class 5. » This page also provides the functionality of selecting which variants and results that are displayed. One could for example sort out all of the information related to variants classified in class 1. The user likes this functionality.

From here the user states that « 1, 4 and 5 goes out to the lab doctor, since 4 and 5are *disease-causing* and a class 1 will never be analyzed again so it is important to get that classification right. » The user wants the report to be uploaded into the database directly without another person having to rewrite the report like it is now with the Swiss lab program. This is an important issue since the «Problem is that people mistype when entering the variant results one by one. Should be possible to load it all directly from the analysis. » The user wants the database of variant classifications to be in-house and all the data should have gone through a quality control.

## Session 2

A recurring issue addressed by the user in this session was that she wanted to see all the variants found in the test-sample to get an initial overview in the first page. The system is designed to limit the number of variants to avoid information overload with the potential huge amount of variants provided by HTS. So the system only shows the variant(s) relevant for the decision making in the current tabs throughout the system. The user states that she would like to see all the variants in the first page to get a gut feeling, but she also states that "since I have not tried the system with HTS data I don't know if it would just be too much." She states further that it would be nice to have the possibility to choose to look over the list of all the gene variants in the test-sample on the first page.

## Work tasks

The user states that "I work with the Lynch genes and analyze seven patients at a time and often I find only one variant in the entire project. It would perhaps take me much longer time to sit and go through the whole workbench analysis seven times. If I could otherwise immediately see that the variants found were normal variants." These normal variants are typical for Norwegians and are recurring in the test-samples. She states that "perhaps this functionality is going to be provided in a new system used prior to the work bench."

## Visual design requirements

The issues related to consistency of button labeling and the font size issues where changed according to the user feedback in session one. In this session there were not identified any issues regarding the labeling of the buttons or font size used in the prototype.

The user states that it should be visualized well if the current user is the first or second person doing the analysis. She further states that now it is marked with in progress/under construction. She states that if the first person selects that a reference is not relevant and the

reference is removed. The second person should have access to the reference since the second person can disagree and think that the reference is relevant.

That the prototype was not adapted to color blind with the primary use of green and red were addressed in this session.

The user states that one problem with the Alamut program they currently use is that all variants with no prior classification are classified as a class 3 variant as a default. The user would much more prefer if there were no default classification.

## Traceability and getting the gene variants into the workbench

The user operates with test number, order number and the patients' national identification number. The user states that the national identification number "is the only thing that is truly unique that can always be used to fetch all the test samples from one patient." But if one can use the national identification number instead of sample-id depends on how secure the system is. She states that if the workbench is used with Sanger then the project number must be registered. To search backwards one has identified a variant and needs to know which project it is linked to. Other reasons might be if one wants to look up prior in database. The user also states that they have started to wright the test results in the VUS-form and asks if it's possible to register the test results in the program and not in the final report.

Contrary to the user in the first user test the user stated that she used the prediction tool(Swift, Polyphen etc.) to some degree, so it could be included in the workbench.

The user states that if she was the second person to do the analysis she would probably take a look in Alamut without looking on the prior evaluation first. She states further that she does not know if others do the same and think that they are not obligated to. This should be evaluated further in other user sessions to evaluate if others prefer this too.

The user states that they enter the name of the lab doctor that they have conferred with regarding the classification in the VUS-form. And states further «this information should probably be included in the system."

The most valued functionalities in the prototype were:

- That the system is going to fit their Workflow.
- Integration of all software's used in the annotation process. Currently the user goes through many programs to find prior classifications.

- The functionality of generating search strings in Google using all different names used on the same variant in the literature.

The two identified tasks that is the most challenging when using the clinical genetic variant classification software:

1. In both the user sessions it was stated that the evaluation of research articles is the most challenging part of the gene annotation process.
2. In both the user sessions there was a focus on how to get the list of variants into the system. And the work done prior to the workbench constructing the list of variants uploaded to the workbench. This work task should be developed to be easy and efficient. Mainly it is the QC of the sequence data, getting rid of *technical artifacts*, check for gene deletions and sorting out the normal gene variants.

# B: The consent form sent out with the survey

Forespørsel om deltakelse i forskningsprosjektet.

## *”Innsamling av brukerkrav til genAP workbench”*

**Bakgrunn og formål**

Dette studiet er en masteroppgave ved Institutt for informatikk ved Universitetet i Oslo. I studien inngår et spørreskjema. Formålet med spørreskjemaet er å kartlegge brukers behov og ønsker for systemet genAP workbench. Resultatene vil bli brukt både til å forbedre programmet og som en del av Jorun Børstings masteroppgave ved UiO i Informatikk - design, bruk og interaksjon.

GenAP workbench er et system for tolkning av sekvensvarianter i klinisk sammenheng, basert på gjeldende SOPer. Programmet eksisterer nå som en første prøveversjon. Spørreskjemaet har et hovedfokus på prosessen ved å evaluere om en referanse er av høy kvalitet, og dermed kan brukes i klassifisering av varianten. Problemstillingen tilknyttet referanser er: «hvordan burde referansedelen av programmet utformes for å ha god brukervennlighet?»

Spørreskjemaet sendes ut til alle som er involvert i tolkning av genetiske varianter ved Avdeling for medisinsk genetikk ved Oslo universitetssykehus, Ullevål.

**Hva innebærer deltakelse i studien?**

Din deltagelse i studien innebærer å besvare et spørreskjema. Skjemaet inneholder 17 spørsmål og tar ca. 30-40 min å fylle ut. Spørsmålene vil omhandle dine meninger om hvordan systemet burde utformes for å støtte ditt arbeid best mulig. I tillegg registrer vi hvilken stillingstittel du har og hvilken enhet du tilhører og hvor lenge du har arbeidet der. Dine svar skrives inn i en Word-fil som leveres som vedlegg på e-post. Deltakelse i studien innebærer at informasjonen du oppgir kan bli brukt i masteroppgaven, evt. også i senere publikasjoner basert på masteroppgaven.

**Hva skjer med informasjonen om deg?**

Alle personopplysninger vil bli behandlet konfidensielt. Din e-post med utfylt spørreskjema sendes til Jorun Børsting (jorubo@ifi.uio.no). Når ditt svar mottas vil Word-filen lagres og e-posten fra deg vil bli slettet. Det beholdes ikke informasjon om din e-postadresse, slik at skjemaet ikke kan kobles til deg. Skjemaene vil deles med Jorun Børstings veiledere ved UiO, Margunn Aanestad og Alma Leora Culén, samt teamet som utvikler genAP workbench. Imidlertid vil ingen av disse ha tilgang til ditt navn eller e-postadresse. Alle svar behandles konfidensielt og din deltakelse vil bli anonymisert i publikasjoner.

Masteroppgaven skal innleveres og studien skal avsluttes 20.12.2014. Datamaterialet anonymiseres fullstendig ved prosjekts slutt ved at stilling, enhet og antall år ansatt ved avdelingen slettes eller grovkategoriseres.

**Frivillig deltakelse**
Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke eller trekke tilbake informasjon som er gitt i spørreskjema uten å oppgi noen grunn. Dersom du senere ønsker å trekke deg eller har spørsmål til studien, kan du kontakte Jorun Børsting.

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Jeg vil innhente ditt samtykke til å delta i studien ved at du besvarer dette spesifikt i første spørsmål i spørreskjema.

Med vennlig hilsen

Jorun Børsting

Kontakt informasjon:

Jorun Børsting

E-post: xxxxxxx

Telefon: xxxxxxx

Du kan også kontakte min veileder:

xxxxxxx

E-post:xxxxxxx

Telefon: xxxxxxx

## C: The Survey

# SPØRRESKJEMA FOR GENAP WORKBENCH

GenAP workbench er et system for tolkning av sekvensvarianter i klinisk sammenheng, basert på gjeldende SOPer. Programmet eksisterer nå som en første prøveversjon, og vi vil veldig gjerne ha tilbakemeldinger på løsninger og forslag til endringer fra deg som mulig bruker av systemet. Derfor har du fått dette spørreskjemaet. Her fokuserer vi på en av de viktigste oppgavene ved variantklassifisering, nemlig vurdering av referanser.

Skjemaet inneholder 17 spørsmål og tar ca. 30-40 min å fylle ut (NB: det er ikke meningen at du skal bruke lang tid på hvert spørsmål. Skriv det som faller deg inn). Vi starter med noen innledende spørsmål, så en forklaring av funksjoner og spørsmål til disse, så mulige tilleggsfunksjoner, og til slutt mulighet for å gi kommentarer på ting som ikke allerede er dekket.

**Informasjon om personvern og anonymitet**
Studien er meldt inn og godkjent av Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS. Skjemaet er laget av Jorun Børsting og Morten Eike. Resultatene vil bli brukt både til å forbedre systemet og som en del av Jorun Børstings masteroppgave ved UiO i Informatikk - design, bruk og interaksjon, evt. også for senere publikasjoner basert på denne. Det er frivillig å delta og du kan når som helst trekke ditt samtykke eller trekke tilbake informasjon som er gitt i spørreskjema uten å oppgi noen grunn.

Alle svar behandles konfidensielt. Svarene dine lagres i form av dette dokumentet, men e-posten fra deg blir slettet og det beholdes ikke informasjon om din e-postadresse eller navn. Svar på spørsmål om stillingstittel, hvilken enhet du tilhører og hvor lenge du har arbeidet ved avdelingen vil også bli slettet etter at masteroppgaven er ferdig sensurert. Svarene vil derfor bli fullstendig anonymisert og vil ikke være mulige å koble til deg i ettertid.

Dersom du senere ønsker å trekke deg eller har spørsmål, kan du kontakte Jorun Børsting (jorubo@ifi.uio.no). For ytterligere informasjon se vedlagt informasjonsskriv.

**Innsending av svar**
Svar fylles ut elektronisk her og sendes som **vedlegg på e-post** til: jorubo@ifi.uio.no.
Bruk emne: «genAP svar», husk IKKE SENSITIV hvis du sender fra en OUS-maskin.

## Innledende spørsmål

1. Har du lest og forstått informasjonen om studien og gir ditt samtykke til å delta i studien? (ja/nei)

2. Om deg:

   Stillingstittel:
   Enhet:
   År/mnd ansatt ved avdelingen:

3. Hvordan finner du frem til relevante referanser for en variant (flere valg mulig)?

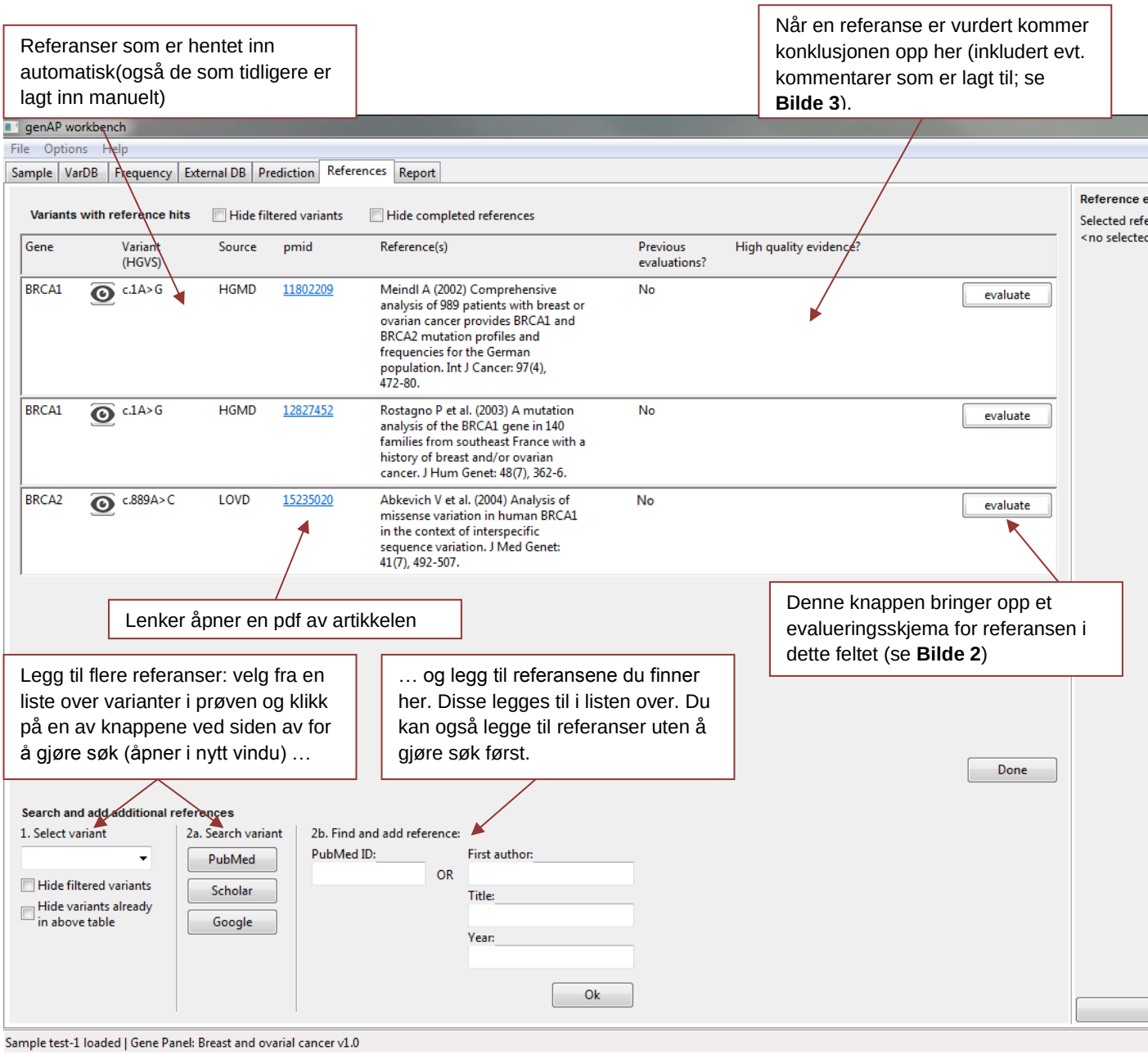
   ☐ Søk på Google
   ☐ Søk på Google Scholar
   ☐ Søk på PubMed
   ☐ I mutasjonsdatabaser (HGMD, LOVD, BIC o.l.)
   ☐ I referanselisten til andre artikler
   ☐ Annet (spesifiser):

4. Dersom du finner flere referanser for en aktuell variant, hvordan velger du hvilken du skal vurdere først?

5. Programmet er laget på engelsk. Er dette til hindring for deg, og ville du foretrekke at det ble oversatt til norsk**?**

## Vise og legge til referanser

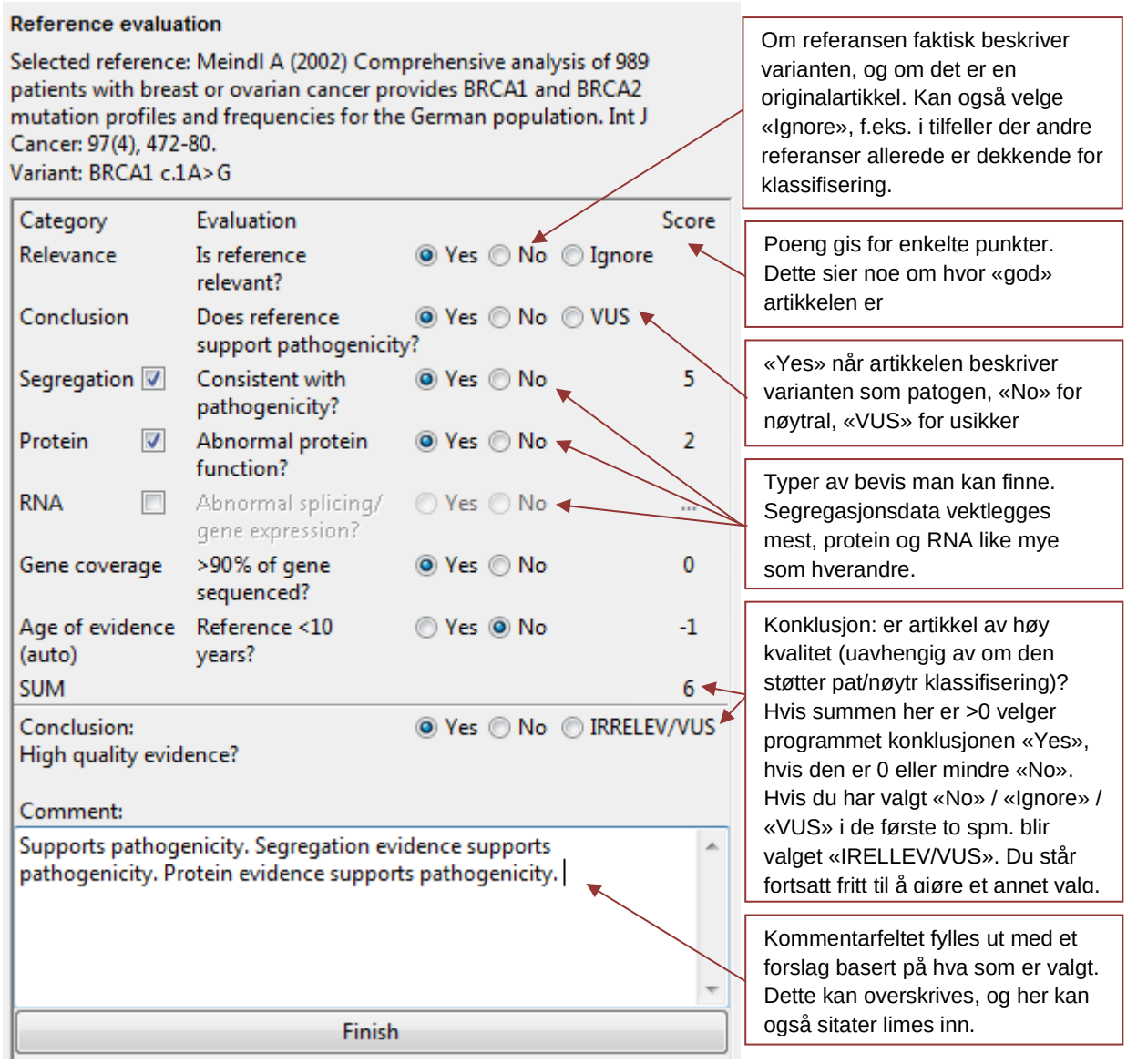
**Bilde 1** under viser det som møter deg når du åpner delen for referansevurdering i programmet. Vi har her tatt utgangspunkt i HTS-data, der alle variantene for flere gener i en prøve er tilgjengelige på samme tid, og der analysen gjøres for én prøve om gangen. Det kan også bli aktuelt med en versjon for andre typer data, der du slår opp for én og én variant. I **Bilde 1** har programmet forhåndssøkt i HGMD Pro, LOVD og interne databaser, og hentet frem de referansene som kan knyttes til varianter som er funnet i den aktuelle prøven (det er disse du ser i listen). Du kan også søke etter og legge til nye referanser (med feltene nederst til venstre).




**Bilde 1:** Referansevurdering i genAP workbench

## Vurderingsskjema for referanser

For hver referanse er det en egen knapp med tittelen «evaluate» (**Bilde 1**). Trykker man på denne får man opp et vurderingsskjema for referansen med veiledende spørsmål på høyre side (se **Bilde 2**). Hensikten er å vurdere om en referanse er av høy kvalitet, og dermed kan brukes i klassifisering av varianten. På bakgrunn av hva man krysser av for genereres en poengsum, der >0 antyder at referansen er av høy kvalitet og 0 eller mindre at den er av lav kvalitet. Dette er kun ment veiledende, og du står fritt til å bedømme referansen annerledes.




**Bilde 2:** Referanseevaluering

Vi har valgt å inkludere segregasjon, protein, RNA, dekningsgrad og alder på referansen som kategorier som kan ha betydning for om en referanse kan sies å være av høy kvalitet (se **Bilde 2**).

6. Virker dette fornuftig? Er det noe annet du synes burde vært med?

7. Burde kategorien «Segregation» være mer differensiert, som å kunne sette inn en evt. LOD-score, eller fylle ut antall generasjoner?

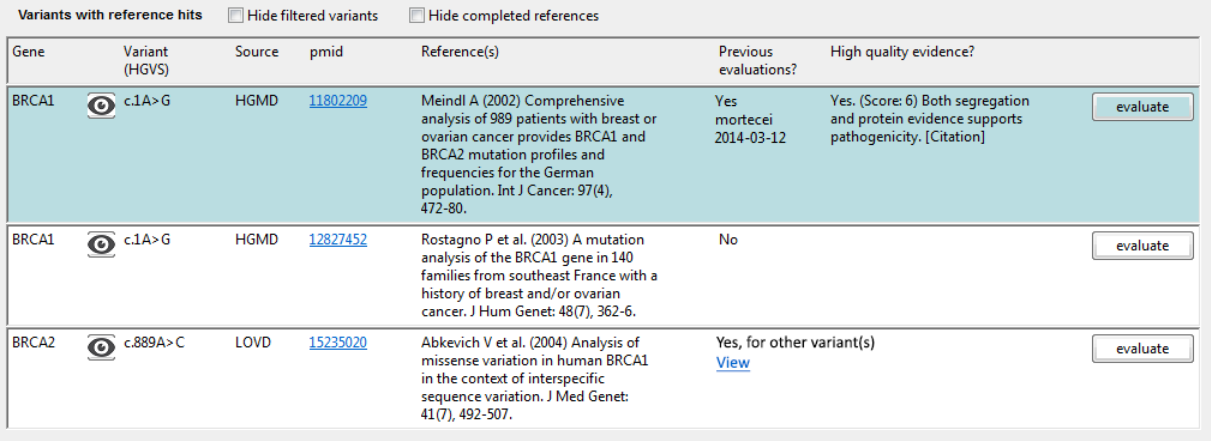
I kommentarfeltet (**Bilde 2**) generes automatisk korte tekststrenger som beskriver hva man har krysset av for. Dette er bare ment som en påminnelse, og det er lagt opp til at du skal redigere dette og legge til egen tekst.

8. Synes du det er nyttig å lime inn viktige sitater fra artikkelen i kommentarfeltet? Hvis ja: Hvordan velger du ut aktuelle sitater?

9. Er det noe annet du synes er viktig å ha med i kommentarfeltet?

## Ferdigvurderte referanser

Når en referanse er ferdig vurdert for en spesifikk variant (av deg eller noen før deg) markeres den med blått og kommentaren fra vurderingen vises i tabellen (**Bilde 3**; i kolonnen «High quality evidence»). Hvis en referanse er vurdert tidligere for en *annen* variant nevnes det også (tredje rad), men her må referansen vurderes på nytt og er derfor ikke markert som ferdig.




**Bilde 3:** Ferdigvurdert referanse (markert med blått). Nederste rad viser en referanse som tidligere er vurdert for en annen variant enn den som er funnet her.

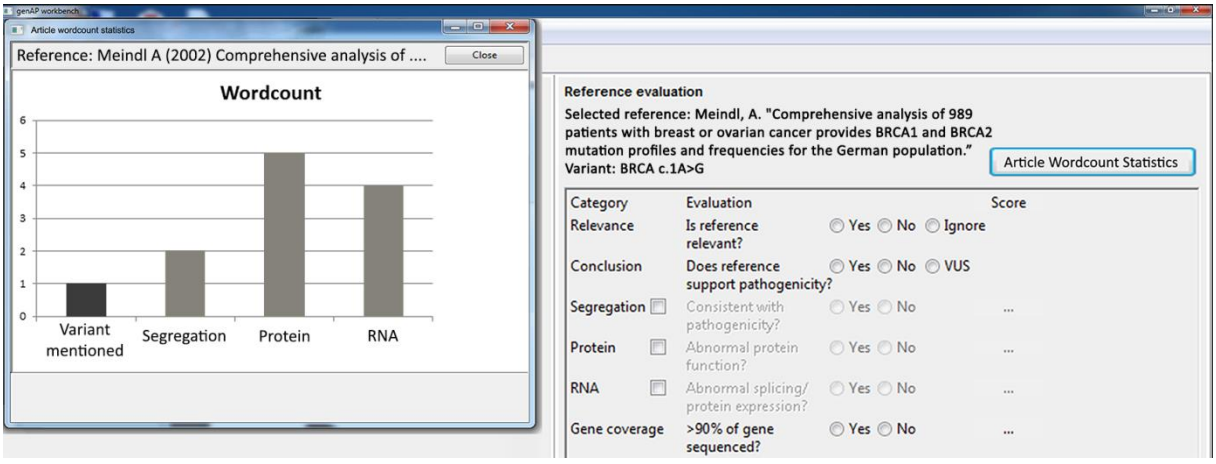
10. Er det annen informasjon om referansene du kunne tenke deg å ha med i oversiktstabellen i **Bilde 3** som ikke er nevnt?

## Mulige tilleggsfunksjoner

Det er mange muligheter for å utvide funksjonaliteten i programmet. Noen av disse har vi beskrevet under.

### Ordtelling

Det er mulig å telle antall ganger et ord nevnes i en artikkel. I **Bilde 4** vises et eksempel på diagram generert ut ifra stikkordene «variantnavn» (svart søyle, på flere alternative måter, inkludert både cDNA og proteinposisjon), segregasjon, protein og RNA i en artikkel. Dette kan du få opp ved å trykke på «Article wordcount statistics»-knappen øverst til høyre.




**Bilde 4:** Ordtelling

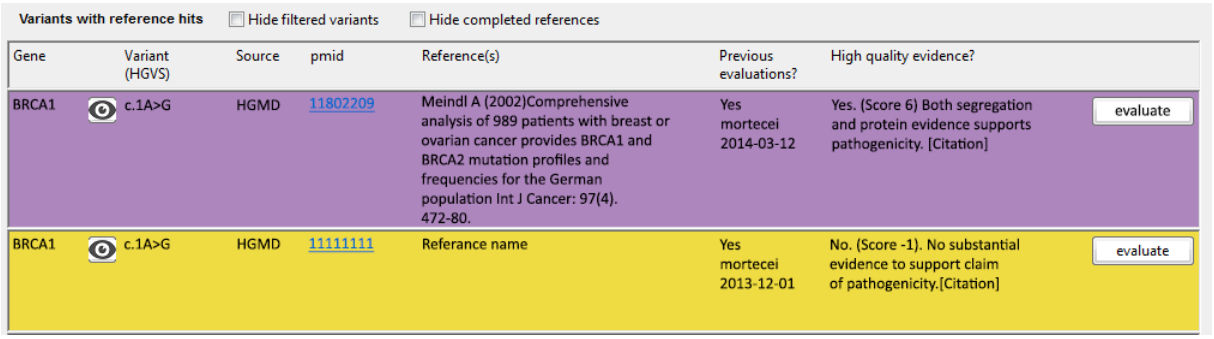
11. Vil denne ordtellingen være nyttig for deg (og hvorfor)?

12. Er det noen andre nøkkelord du synes er relevant å søke etter på denne måten?

13. I tilfeller der man har flere referanser om samme variant: er det ønskelig å sammenligne ordtellingene i disse? Hvordan vil du eventuelt bruke dette?

### Alternativ fargekoding av vurderte referanser

Programmet sånn det er nå markerer referanser som allerede er vurdert i blått (**Bilde 4**). Alternativt kunne man tenke seg andre/flere fargekoder for å gi mer informasjon. I **Bilde 5** vises én slik mulighet, der en referanse som er vurdert som god er farget lilla, mens en som er vurdert som dårlig er farget gul. En variasjon av dette kan være å farge ferdigvurderte referanser blått som standard, men der du som bruker kan gi utvalgte referanser blant disse en annen farge, f.eks. lilla for en spesielt god referanse, og gult for en referanse som har ting som gjør at man bør være spesielt kritisk. Det er også mange andre muligheter, f.eks. å farge referanser røde eller grønne basert på om de hevder at varianten er hhv. patogen eller nøytral.




**Bilde 5:** Alternativ fargekoding av ferdigvurderte artikler. Lilla: referanse av høy kvalitet; gul: referanse av dårlig kvalitet.

14. Hva synes du om å merke vurderte referanser med farger på denne måten, og hvordan synes du det bør gjøres?

15. Kunne du tenke deg å bruke fargekoding på en annen måte enn forslagene over? Evt. bruke noe annet enn fargekoding for å få frem det samme?

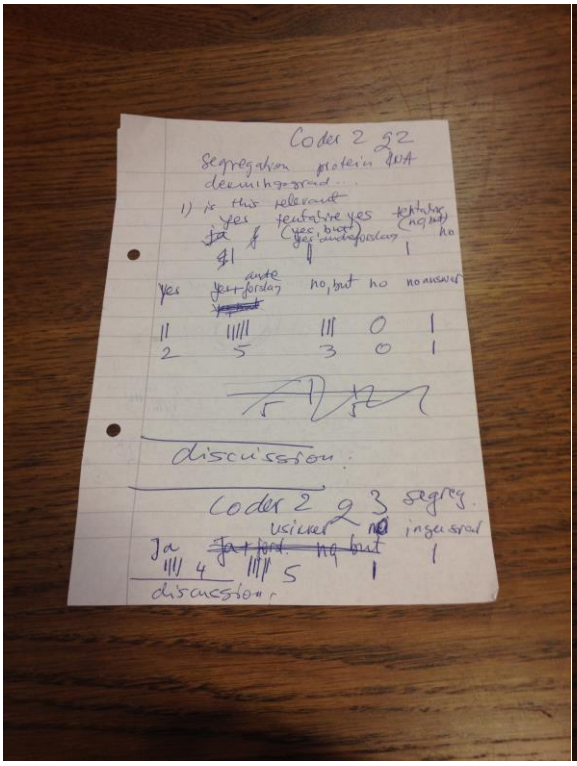
## Annet

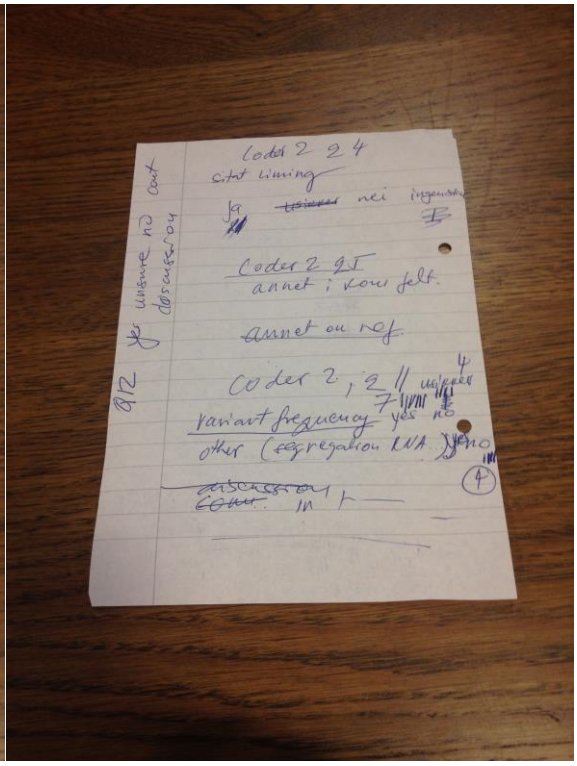
16. Dekker programmet som det er beskrevet hvordan du i dag vurderer en referanse? Hvis nei, hva synes du mangler?

17. Har du noen andre kommentarer eller forslag til alternative løsninger for referansevurdering som ikke er dekket av dette spørreskjemaet?

# Tusen takk for hjelpen!

# D: Pictures from the content analysis



# E: Pictures from the data analysis