IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 13, 2021, accepted February 15, 2021, date of publication March 2, 2021, date of current version March 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063468

The Role of a Facilitator in Co-Design Applications for Exploratory Analysis in Domains of High Complexity: The Case of MAHiCGO

TALAR B. YACOUBIAN[1], DENA A. AL-THANI[2], AND MICHAËL AUPETIT[3]
[1]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[2]Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[3]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

Corresponding author: Talar B. Yacoubian (talar.yacoubian@gmail.com)

ABSTRACT **Background**: Robust and useful tools for exploratory analysis in biosciences are still lagging behind the size and complexity of the biological datasets produced since the completion of the human genome in 2000. A possible reason is that developers are unlikely to understand domain and case-specific requirements of existing research questions. **Methods**: We formed a design team comprising a visualization expert, a Human Computer Interaction (HCI) researcher, bioinformatics domain experts, and the Principal Investigator (PI) as a 'facilitator' filling the communication gap between them. We implemented co-design methodology. **Results**: We identified the need for an interactive visual analytic tool for exploratory analysis of biological data. We describe the process of developing MAHiCGO, a novel tool for the simultaneous visualization of MA Gene Expression data, Hi-C data and Gene Ontology (GO) information in an interactive manner for the exploratory analysis of biological data. **Conclusion**: The key finding of this research to include a facilitator role in the co-design is useful in the evolving fields of design research and of bioinformatics, merging computational sciences with biosciences. The findings support in understanding new functional roles in the field of design, in particular design of computer applications in highly complex domains such as bioinformatics, and for highly complex tasks such as exploratory tasks. The findings also stress on the key role of visualization to expand user cognitive capabilities, and of co-design for constant engagement of domain experts in the creation process.

INDEX TERMS Co-design, applications, HCI, visualization, BioVis, visualization for exploration, interactive visualization, cross-domain collaboration, bioinformatics, MAHiCGO.

I. INTRODUCTION

A. OVERVIEW

John Tukey characterizes Exploratory Data Analysis (EDA) by visualization, flexibility, and search for simplicity. EDA emphasizes on the discovery of the unexpected by revealing patterns and features of the data through visualization [43]. Despite this description of EDA since the seventies, visualization is still widely seen as a tool to represent results rather than a tool to support human cognitive abilities in exploratory analysis [3].

For a design team to co-create, or to co-design computer applications, the following functions are required: **to understand user needs, to design prototypes, to develop the designed prototypes, to give and to receive feedback through evaluations [20], [45], [47]**. It is ideal for these functions to be carried out by referring to the literature and research for the most appropriate theoretical and practical methodologies. This requires **Human Computer Interaction (HCI) expertise [18], [28]**. While for cognitively complex prototypes, **visualization expertise** could add great value to the co-design process [13], [23], [27], [49].

Distinct individuals in a co-design team could carry out each of the different functional roles identified thus far. Alternatively, the same individual can carry out more than one function. For example, the same person could carry out design and development functions. Understanding user needs and identifying best practices through research could be the role of an HCI researcher.

Mirel *et al.* [22] have argued that one reason for the fact that sufficiently robust and useful tools for exploratory analysis in bioinformatics is in its infancy is because developers do

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti.

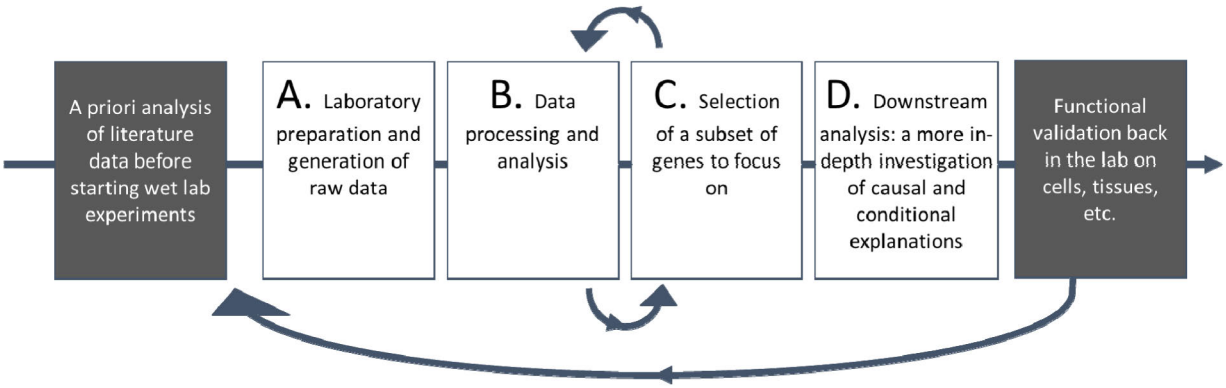

FIGURE 1. Understanding the user workflow. The user journey starts with laboratory preparation and generation of raw data. The following two stages, data processing, analysis and selection of a subset of genes to focus on, is where MAHiCGO is designed to help the users in preparation of the final stage of downstream analysis. Section IV describes the first prototype of the visual analytic tool MAHiCGO in detail.

not fully understand domain and case-specific requirements of existing research questions [22]. Before the co-design era, the functional role of 'understanding user needs' has been generally carried out by the HCI researchers. In co-design research, designers, developers and HCI experts carry the role out collaboratively [28], [38].

Co-design has come to shift the design process from designing for users to designing with users. Co-design requires the whole team to create throughout the span of the design process [28], [38].

For a user domain of high complexity such as bioinformatics [22]–[26], [49] neither the designer, nor the developer, nor the HCI researcher will necessarily have a solid or even a general understanding of the domain. Similarly, domain experts might not have the required level of creativity to design [24], [28]. It might require extensive time and effort to bring domain experts to a level to acquire the desired creativity to design, if at all possible.

B. MOTIVATION AND AIM

To tackle the challenge of understanding user needs for the highly complex bioinformatics and biological visualization (biovis) domains, we followed a co-design methodology. To facilitate understanding and translation of needs, we assigned a unique individual with enough knowledge and expertise in all three fields of bioinformatics, human-centered research, and design. The individual is a graduate student in Data Analytics in Health Management, and had covered introductory courses in bioinformatics. Throughout the paper we will refer to this functional role as **the Principal Investigator (PI).**

With a team of an HCI researcher, a visualization expert and the PI, we worked with domain experts (users) through a co-design process.

Through background research and interviews with the domain experts participating in this study (detailed in Section III) the team was able to narrow the wider field of bioinformatics to focus on RNA-Seq research. We shaped an understanding of the conventional workflow of genomic data analysis comprising four broad stages (Fig. 1). Stage 'A' includes laboratory preparation and generation of raw bioinformatics data generally in conventionally agreed formats. Stage 'B' comprises analysis of the generated data in some sort of visualization or through the use of traditional databases. Stage 'C' includes selection of a subset of the data based on inferences from the analysis. Finally, Stage 'D' includes downstream analysis, a more in-depth investigation of causal relationships in biological context, including any or a combination of Enrichment Analysis, Network Analysis, Motif Analysis, and others. Some interviewees in this research have identified a stage before 'A', the a priori analysis of literature data before starting wet lab experiments to increase the likelihood of the experiment success. Others have identified a stage after 'D', the functional validation on cells, tissues, etc. back in the laboratory. Most users of this research generally view stages 'B' and 'C' as part of a single stage where innovative ways of visualization and representation of data can be very useful and can lead to potentially valuable discoveries through EDA (Fig. 1) [30], [31], [50] (Personal Communications, 2019-2020).

In their 2019 comprehensive literature survey to compile ''Tasks, Techniques, and Tools for Genomic Data Visualization'', Nusrat *et al.* [24] highlight how complex information from different sources need to be integrated into visualization in order to understand and interpret genomic information and the corresponding complex biological systems. These can be information about the abundance of proteins, or about chemical modifications, or relating to 3D structures of chromosomes. ''The regulation of gene activity'', the authors write, ''is controlled by the presence or absence of particular regulatory proteins, chemical modification of parts of the DNA, and the 3D structure of chromosomes, all of which are changing depending on environmental and other factors.''

An example the authors give in explaining the importance of putting multiple data types into context is: evaluating if a certain physical gene mutation has a significant functional

impact by simultaneously visualizing the corresponding gene annotations of the mutated and the non-mutated genes [24].

We identified a need to support interactive exploratory analysis of gene expression from two biological samples under study - for example a random sample and a control group. The specific application that has been identified is complex because it involves RNA-Seq differential expression data (sometimes referred to as MA data) [20], Hi-C spatial data of chromosomes [10], [13], and Gene Ontology (GO) information [1], [39] for exploratory analysis of the significantly expressed genes between the two groups under study.

As an outcome of this co-design process, we developed a first prototype of the MAHiCGO visual analytic tool. The developer co-designing and implementing user needs is a visualization expert. The assignment supported the complex task of exploratory analysis in a highly complex field by expanding human cognitive abilities through visualization. Moreover, we reflected on the co-design process and concluded that the role of the PI, having the characteristics of a facilitator elaborated in section V, is important for computer applications for complex tasks like EDA and in complex domains like bioinformatics.

We structured the remaining of the paper as follows. Section II provides an overview of HCI research in biological visualization. Section III describes the methodology through which we conducted this study. Section IV discusses the results and research findings. Section V concludes the paper by presenting general recommendations and future work considerations.

II. BACKGROUND AND LITERATURE REVIEW

The following section starts with an overview of the historical development of the fields of bioinformatics and biovis with a specific focus on the three visualizations used in MAHiCGO: MA, Hi-C, and GO. The overview is followed by a presentation of the challenges associated with a wide variety of biovis tools. The section concludes with an overview of HCI design in biovis and the drivers for choosing co-design as the theoretical background of this study.

A. THE COMPLEXITY OF VISUALIZATION OF BIOLOGICAL DATA (BIOVIS)

Lander *et al.* [22] characterize the 20th century by four phases that have propelled biological sciences to new levels. The first phase established the chromosomes as the cellular basis for heredity. The second phase defined the DNA double helix as the molecular basis for heredity. The third phase uncovered the mechanism by which cells read the information in genes. The last phase saw the birth of genomics – the drive to decipher the first genes and genomes. With the announcement of the completion of the human genome in the year 2000, the field of biology was faced by a sequenced genome eight times as large as the sum of all previously sequenced genomes [17], [26].

In his ''Rise and Demise of Bioinformatics? Promise and Progress'', Ouzounis describes how the field of bioinformatics has gone through the stages of infancy, adolescence and adulthood between the years 1996 and 2011 with persistent problems of data organization, accessibility and interpretation [26]. It forecasts that by 2020s the field of bioinformatics and computational biology will emerge as a distinct discipline with redefined scope as computational sciences continue to merge with biosciences and biomedicine [26].

Humans can be good in making sense and memorizing a certain volume of information, but as the data gets larger, interaction seems necessary in sense-making efforts including scanning, recognizing patterns and identifying relationships and interpreting the data. Interaction supports the analytical aspect of Information Visualization and expands human cognitive capabilities even more than static visualization [23], [42], [49].

Over the last decades, biological visualization has matured exponentially to produce an array of diverse visualization tools but has not been able to match the increase in size and complexity of biological datasets produced. The concept of developing the corresponding tools have been emphasized in recent years, but the pace of development of usability has been slower than competing priorities because ''usability is usually less rewarded in science than is research on new methods'' [25].

The biological visualizations specific to this study are the MA plot, the HI-C data visualization, and GO information visualizations.

1) MA PLOTS

MA plots are visualizations available in widely used tools for differential gene expression analysis, such as edgeR and DESeq2, to visualize genes expressed differently between two samples under study - normal versus mutant, or wild type versus knock out. For studies with more than two samples, the researcher develops an MA plot for every pair of samples [20].

MA plots have evolved to well-developed representations of gene expression data; but researchers mostly utilize them as static, standalone outputs of gene expression RNA-Seq experiments, which are widely used for understanding gene expression data between two samples.

In the MA plot representing thousands of significantly expressed genes from a biological experiment, if you know which genes (or set of genes) you are looking for based on previous knowledge, it might be easy to visualize the specific gene on the plot and examine the corresponding intensity of differential expression for the specific gene (or set of genes). Nonetheless, the representation of thousands of genes might not prove useful or intuitive for researchers aiming at exploratory analysis of the data. In our approach to develop the MAHiCGO prototype, we propose an interactive MA plot, which allows the user to select genes of interest, to filter genes based on the p-value, and to visualize the selected genes in the other two plots in a simultaneous and interactive manner.

2) Hi-C PLOTS

Hi-C data, a transformation from 3C, 4C, and 5C Chromosome Conformation Capture assays, represents the spatial arrangement and interaction of chromatin material in the three dimensional space of a nucleus for the whole genome. Unlike other sequencing techniques (e.g. RNA-Seq or ChIP-Seq) that result in one-dimensional vectors of the genome, Hi-C data comprises a two dimensional matrix, with the data points representing interactions of each pair of genomic loci [10], [13]. The tools developed as part of this research utilizes the visualization of Hi-C data on a diagonally symmetric heat-map that usually represents strong correlation of genomic regions of close proximity. The heat-maps in MAHiCGO are interactive allowing selection of pairs of genes and visualization of the selected genes in the other two plots MA and GO.

3) GO PLOTS

As with any ontology, Gene Ontology encodes biological information in a hierarchical, general-to-specific way. The standardized representation of the biological information has rendered GO a highly useful tool to interpret, for example, the location or function of an over-expressed or an under-expressed gene from a high-throughput sequencing experiment, or to look for genes of similar characteristics [5]. This body of ontology knowledge of biological systems is how the GO project classifies our biological knowledge of today. GO diagrams represent the wealth of information about biological processes as nodes and edges, the latter representing the relationships between the nodes. Nodes are GO terms having at a minimum: a unique identifier (e.g. GO:004428), a label (e.g. 'small molecule metabolic process'), and a definition; but can also contain additional information [1], [39].

In addition to the collection of biological terms and their relationships, GO includes a wealth of knowledge of linking genes (or gene products such as RNA) to GO terms. This process is called gene annotation and it is a statement about the most up-to-date function of a gene in biology by associating it to a GO term and by transitivity principle to all its parent terms. If the role of a gene product is unannotated in GO, it means its role in biology is still unknown [39]. Networks are widely used modes of visualizing groups of genes with corresponding GO terms.

In an era of a multitude of visualization packages, biologists still have user-interface challenges finding tools difficult to learn or challenging to meet exact requirements and therefore still requiring custom solutions or the need to learn at least one programming language [25].

B. THE CHALLENGES ASSOCIATED WITH A WIDE VARIETY OF DISCONNECTED TOOLS IN BIOVIS

R is a programming language used for the development of most of the packages in Bioconductor, an open source and open development environment that provides tools for the analysis and comprehension of high-throughput biological data. The first packages for analysis and interpretation of biological data generally lacked any Graphical User Interface (GUI). Later, and as frameworks developed to provide GUI in command-line R packages and elsewhere, developers started to incorporate packages like SeqGrapheR and OLINgui to provide complementary GUI to existing biovis tools. Developers targeted most sequencing and further downstream analyses tools, specifically the ones without advanced GUIs, to bioinformaticians due to the required computational skills. Tools specific for biologists without a solid knowledge of bioinformatics or computational sciences are limited in scope and hard to generalize [50].

Many of these biovis tools have remained separate from conventional computational analysis tools and programming languages. Batch & Elmqvist argue this being one of the main three reasons behind the fact that visualization tools are not widely used as part of the analysis process [3]. Similarly, Nusrat *et al.* highlight the need to improve integration among different tools to support researchers or users in a more holistic understanding of the matter at hand and to allow for easier transition from one analyzing tool or environment to another [24].

Despite the abundance of visualization tools to represent information conveyed by genomic data, many argue that the large variety presents a dilemma to end-users to find domain or task-specific tools [30], [31]. Another problem most biologists face while working with these tools is the heavy reliance on coding and programming expertise which most biologists might need to exert extra effort, in addition to their existing skills and expertise, to acquire [24], [50].

Despite the relative advancement of the GUI to visualize MA, HI-C, and GO, none of the available tools support a multi-way, simultaneous and interactive visualization of the biological data underlying MA, Hi-C and GO thus making the process very cumbersome for biologists to correlate these data and make potentially important discovery in the field of biology (Fig. 2).

By copying and pasting long lists of genes from one software to another, biologists try to make biological sense and inferences from the data at hand. Tools have recently started to incorporate interactivity in different visualizations. Examples include HiGlass[1] and AEGIS.[2]

In line with this study's underlying assumption of domains of high complexity and to ease the flow of the text, we have included domain-specific technical information relating to MAHiCGO in the appendices in case the reader requests further understanding. Appendix A presents state-of-the-art biovis tools and practices gathered through interviews with domain experts followed by verifying identified gaps through literature review.

[1] http://higlass.io
[2] http://aegis.stanford.edu/

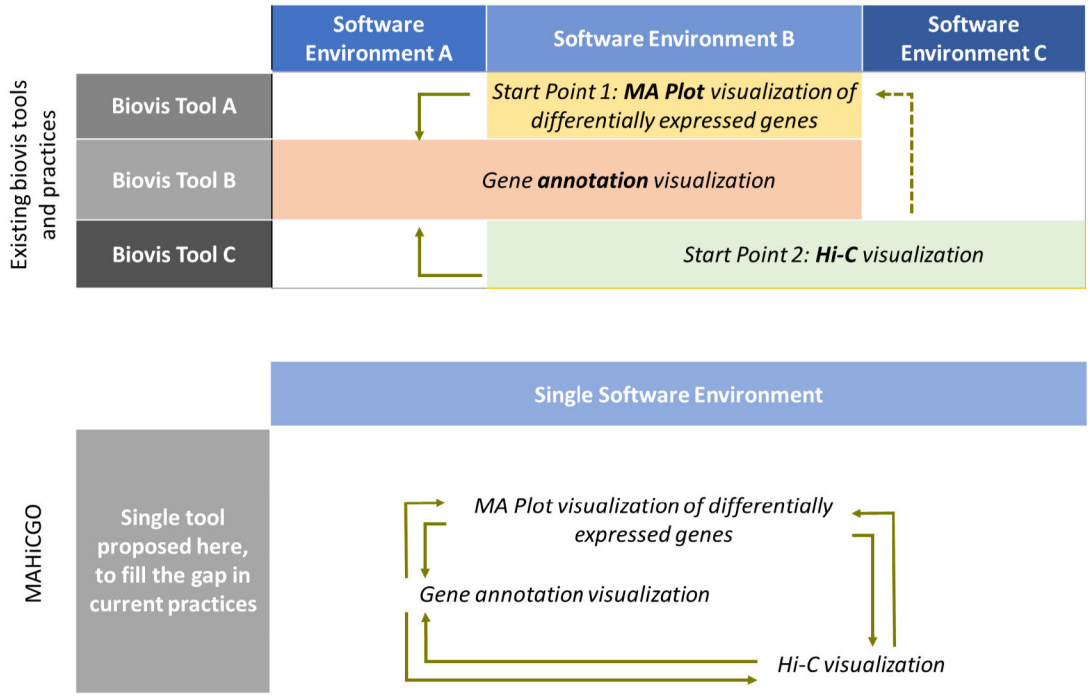

FIGURE 2. Overview of the status quo of existing biovis tools in the market (top). For the analysis of MA, Hi-C and GO information, there is a gap in integration of different existing tools in different software environments. MAHiCGO is a novel tool that will fill the gap in the existing practices, interactively combining all three visualizations in a single biovis exploratory tool for analysis (bottom).

C. HCI DESIGN IN BIOVIS: METHODOLOGIES TO DEVELOP BIOVIS TOOLS

Interaction design hackathons, referred to as design jams, exist to present means to identify problems, to share knowledge and to define community goals for interaction design purposes [40].

User experience (UX) and interaction design hackathons exist but are not widely adopted in biosciences communities. Thomer *et al.* argue that this is not out of disinterest but rather a lack of resources, initiatives and cross-domain collaborations [40]. In their 2018 work, Batch & Elmqvist [3] further underscore the observation that visualization experts need to be taking the initiative to highlight the importance and promote the use of visualization and interactive GUI for exploratory analysis to end-users in general and data analysts in particular [3].

The evolution in design research from a user-centered approach to co-designing is changing the roles of the designer, the researcher and the person formerly known as the 'user'. ''Sometimes 'users' can play co-creating roles throughout the design process, i.e. become co-designers, but not always. The person who will eventually be served through the design process is given the position of 'expert of his/her experience', and plays a large role in knowledge development, idea generation and concept development''. The level of involvement in co-creation of this person, previously the 'user', now the 'expert of his/her experience' as referred to by Sanders & Stappers, depends on the level of expertise, passions and creativity [28]. We refer to them as domain experts in this research.

Scientific software developed by domain scientists with little knowledge in user interface (UI) or UX design, or by software engineers with broad system knowledge but little knowledge in domain needs and visualization requirements, might not fulfill the necessary requirements for a GUI with high usability [34]. Thomer *et al.* [40] have highlighted that not involving the right expertise in the development process of a software with rich and useful GUI could hinder the whole development process.

Co-design, by definition, refers to the process of engaging all design team members to apply mutual empowerment and collective creativity from the beginning and throughout the lifecycle of the design process to make sure user needs are met. The notion of designing for the experience rather than just the product has later been introduced [28]. The focus of our design has been to support for exploration. A typical process in co-design starts with a session where the technical expert/ visualization designer explains the application ranges and possibilities of the domain of visualization thus raising awareness of the field to the domain expert; and it is followed by a thorough session of understanding current practices delivered by the domain expert. These first two phases allow both parties to expand their knowledge about each other's areas. The sessions are followed by long, and at times iterative, discussions about the potential aspiration and requirements of the domain expert which are later

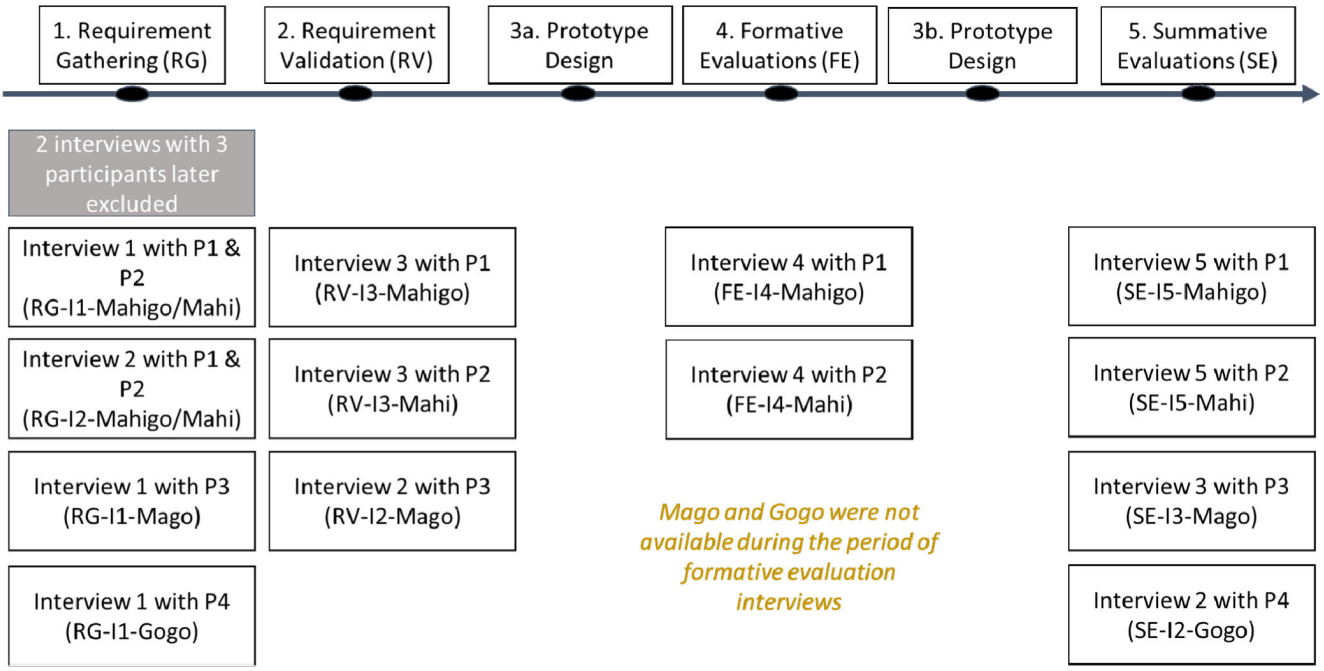

FIGURE 3. The five phases of the research: Requirement Gathering, Requirement Validation, Prototype Design, Formative Evaluations, and Summative Evaluations.

refined and shortlisted by the technical expert/ visualization designer [34].

The most common way to engage all co-design team members in co-creation and co-design is through design workshops [13], [20], [47]. But there could be different ways for the process [13]. In line with the principle of mutual empowerment of co-design team members to contribute to the design process, we have followed a methodology of mutual empowerment for design throughout all the stages described in Section III. The technical (HCI and visualization) experts introduced the wide spectrum of support interactive visualization could provide to the EDA process and demonstrated state-of-the-art visualization tools including Shiny from R [33]. The domain experts correspondingly identified the different design requirements of the different features of MAHiCGO in iterative discussions and validation sessions.

The field of design research is evolving [28], [38]. Sanders & Stappers highlight that as new roles shape, the future of co-design language will need to fill the gap in the functional role for communication between the design team, the various levels of 'user groups' and other stakeholders [28].

III. METHODOLOGY

As a typical co-design process, the design journey in this research study started with an iterative involvement of domain experts who will eventually be served by the designed prototype of the visual analytic tool MAHiCGO. The process included five phases (Fig. 3) and four domain experts (participants) referred to as Mahigo (P1), Mahi (P2), Mago (P3) and Gogo (P4). This section elaborates on the selection of participants and the procedure of the study.

A. PARTICIPANTS

The PI, together with an HCI researcher and a visualization design expert, started the recruitment process with two domain experts and recruited additional participants through Peer Esteem Snowballing Technique (PEST) [4]. PEST starts with an objective to identify a number of informants who will provide unbiased initial nominations based on which further nominations will be generated in a series of waves. Christopoulos argues that to target a small number of expert population, this technique provides a level of confidence in the representativeness of the sample population [4].

Through PEST, seven potential candidates were identified. The inclusion criteria used for the selection process were twofold. a) Domain expert was to be a biologist or bioinformatician experienced in and/or currently conducting research on analysis of RNA-Seq data. This resulted from the initial informal conversations with experts in the field showing that RNA-Seq seems to be the most widely studied methodology in genomic research in general and at the research complex where the interviews were being carried out in particular. And b) domain expert to be available in Doha, Qatar for face-to-face interviews or laboratory evaluation sessions. Following the above inclusion criteria, we recruited four participants from an initial pool of seven.

The four participants with the profiles presented in Table 1 are referred to as Mahigo, Mahi, Mago and Gogo

TABLE 1. Profile of participants.

Domain Expert	Qualification	Tenure	Current occupation
Mahigo (P1)	PhD in Molecular Biology	9 years	Assistant Professor, current research heavily rely on MA Data, Hi-C Data and GO
Mahi (P2)	MSc in Molecular Cell Biology	1 year	PhD Student, current research heavily rely on Hi-C and MA Data
Mago (P3)	PhD in Genetics	15 years	Principal Investigator/ Director of a Research Team, current research relies on MA and GO Data
Gogo (P4)	PhD in Animal Biology	8 years	Postdoctoral Fellow, current work heavily involves GO analysis

based on their current research focus and their expertise with the selected visualizations – MA, Hi-C, and GO.

In line with Kandel *et al.*'s [12] classification of data analysts, all four described themselves as mostly **Scripters/Hackers.** Kandel *et al.* define hackers as users of a mix of programming languages and visualization environments for analysis. Scripters are those who, even if not performing their own ETL (Extract, Transform and Load), use an analytical environment (e.g. R) with complex statistical modeling [12].

Because the profession or expertise in bioinformatics/biology of the user is critical to the evaluation exercise, it is common to have small number of participants in similar studies where it is hard to find a large number of highly specialized participants with the suggested expertise [4], [20].

The fact that all participants in the study are Hackers/Scripters is a pure coincidence of availability of participants in line with the inclusion criteria and is not intentional as part of the methodology. Nonetheless, even biologists with knowledge and experience to script would be willing to be provided with tools that will create an easier environment for analysis, exploration and discovery [31].

B. PROCEDURE

The procedure to conduct this research included five phases (Fig. 3): Requirement Gathering, Requirement Validation, Prototype Design, Formative Evaluations, and Summative Evaluations.

All requirement gathering and evaluation sessions were carried out as open-ended, think-aloud [7], semi-structured interviews. We used questionnaires to evaluate usability (see Appendix B). The sessions took place in participants' workplaces and therefore the setting was as close to a real work setting as possible. The co-design team members observed domain experts performing predefined list of tasks. They used the structured observation method [47] in a real setting.

All sessions were video or audio recorded upon the participant's permission and upon the mutual consent that once the study is completed all recorded information will be destroyed. We used the collected data for transcribing and analysis purposes. The video recordings were mainly that of computer screens and participants' hands interacting with the computers.

1) PHASE 1– REQUIREMENT GATHERING

This phase included the seven potential candidates recruited by the PI with the support of the visualization and HCI experts through PEST. We interviewed the first three domain experts who were not selected to be part of the study following the inclusion criteria. The Requirement Gathering phase continued with four interviews with the four domain experts included as part of this research.

The PI facilitated four requirement gathering interviews RG-I1-Mahigo/Mahi, RG-I2-Mahigo/Mahi, RG-I1-Mago, and RG-I1-Gogo (Fig. 3) with two-fold aims. The first aim was for the designer to raise awareness of the possibilities that could be achieved by potential visualization tools. Demonstrations included different available tools and the different products with a specific focus on interactivity capabilities. The second aim was for the domain experts to describe a typical workflow of RNA-Seq analysis and identify, together with the visualization expert, potential areas of support through visualization (Fig. 1). To gain a solid understanding of the user journey in analyzing RNA-Seq data and identifying areas where interactive visualization could be useful, the participants were asked to describe the way they perform RNA-Seq analysis and to demonstrate their work in programming environment they use, and to list the tools they currently use.

The four requirement gathering interviews prepared all members of the co-design team for mutual empowerment to design. Since domain experts, biologists and bioinformaticians, might have limited view of the possibilities available by the visualization domain, and since visualization designers might have limited awareness about the domain-specific biological knowledge, these two aims complemented each other by providing equal opportunities for both groups to learn from each other. Detailed questions asked during the interviews can be found in Appendix C.

2) PHASE 2 – REQUIREMENT VALIDATION

The PI facilitated and conducted three requirement validation interviews RV-I3-Mahigo, RV-I3-Mahi, and RV-I2-Mago (Fig. 3) due to availability considerations of domain experts. The visualization expert, who will be developing the prototype, and the HCI expert actively listed and supported in presenting solutions to the teams. Following every session, the

PI, HCI expert and the visualization expert conducted a thorough analysis of the information gathered. They transcribed the minutes of the initial requirement gathering sessions and shared them with the domain experts for transparency and validity purposes. They incorporated any feedback from further evaluations in the design process.

It is important to highlight that throughout the requirement gathering and the requirement validation phases, PI's role was key in realizing a smooth collaboration between all members of the co-design team. The PI facilitated understanding user requirements in the language spoken by biologists/bioinformaticians. The PI also supported in translating the design requirements to the designer, and the designer's and HCI expert's comments into a language understandable by the biologists. The PI also carried out the required desk research in all three fields to fill the knowledge gap in each other's domains, and confirm the identified gaps in design through a study of the state-of-the-art visualization tools in biology.

Detailed information gathered from the interviews and desk research relating to the conventional biological data analysis workflow and the state-of-the-art biovis tools can be found in Appendix A.

3) PHASE 3 – PROTOTYPE DESIGN

Following phases 1 and 2, the visualization expert conducted the technical development of the first prototype. The HCI expert conducted the UX design. The PI participated in the UX design and development process. After the formative evaluations described in Phase 4, the design continued for an enhanced prototype of the tool.

As one of the important features of co-design, and to ensure optimal engagement in early prototyping, the datasets were provided and were pre-processed by one of the domain-expert users (Mahi). Pre-processing the data in advance of the evaluation sessions also rendered the sessions more effective in terms of focusing on the objective of studying the interactive visualization portion of data analysis and exploration of results. This has been highlighted by Saraiya *et al.* [31] who underscore that biologists' goal is to identify and understand complex interactions among genes and conditions and correspondingly highlight the importance of studying the interactive visualization portion of data analysis. In doing so, their focus in the study is on evaluating the success of the visualization portion of datasets that have been pre-processed, normalized, pre-filtered and converted into the required formats [31].

The following datasets were provided by Mahi to be used during the development phase.

- Two cooler datasets: One for GM12878 cell line and another for H1hESC cell line.[3]
- Differential gene expression between GM12878 and H1-hESC cell lines (using DESeq2 package).
- Reference gene annotation file for Human Genome 19 (hg19) genome (gtf format) (from UCSC database[4]).

4) PHASE 4 – FORMATIVE EVALUATIONS

To engage with the users throughout the prototype design process, the PI facilitated and conducted two formative evaluations including only Mahigo and Mahi due to availability considerations. We have resorted to the User Experience (UE) evaluation methodology [16] in our study and we have formally reported the systematic process in Chapter IV. The visualization expert and HCI expert were active listeners during these sessions.

The formative evaluations FE-I4-Mahigo and FE-I4-Mahi (Fig. 3) were qualitative in nature with each of the end users having minimal to none training of the first version of the prototype being evaluated. During these formative evaluation sessions, both domain experts and the rest of the co-design team gathered further insights about the usability and learnability of the prototype. The team detected overlooked requirements, flaws and deficiencies as domain experts carrying out the evaluations came across the prototype for the first time.

After the completion of formative evaluations, the prototype design continued. The visualization expert conducting the developer role incorporated user comments in the design process. Informal discussions continued to take place between the domain experts and the rest of the co-design team about the progress in designing the prototype. The purpose of these iterative collaborations in a co-design approach is to ensure that the developer understands exact user needs and work practices, and domain experts are comfortable with the prototype the designers produce commenting on flaws and faults throughout the design process.

During FE-I4-Mahigo and FE-I4-Mahi (Fig. 3), the team asked the domain experts to identify tasks to be used in the subsequent summative evaluations. The purpose of this exercise was to trigger the thought process of the domain experts for exploratory analysis exercising some pressure to understand and explain their own EDA process. Due to the exploratory nature of the tool, it has proven difficult for end-users to identify specific tasks at this stage.

5) PHASE 5 – SUMMATIVE EVALUATIONS

Finally, the PI facilitated and conducted four summative evaluations carried out individually with the four participants. We refer to them as SE-I5-Mahigo, SE-I5-Mahi, SE-I3-Mago, and SE-I2-Gogo (Fig. 3).

Through a thorough literature review, further discussions with users, and a better understanding and insight of the analysis and the exploratory workflow of the data under study, the PI identified the following tasks prior to the summative evaluations.

- T1: Find an interesting set of differentially expressed genes describing the reasons for the choice; identify

[3] https://github.com/mirnylab/cooler

[4] https://genome.ucsc.edu/cgi-bin/hgTables

where the genes are on the chromosomes and their interactivity with other genes. What conclusions can you draw? What can you say about their functional GO categories?
- T2: Starting from the Hi-C dataset, select a set of interesting genes describing the reason for selection. Discuss your observations about their differential expression and functional GO categories.

At the start of each session unlike the formative evaluations, the team provided a brief explanation of the functionalities of the prototype, the latest enhancements, and information about the datasets used. To spur the thought processes of domain experts and ensure optimal engagement level, the team asked domain experts to write down a task they would perform with the given dataset. Now that the prototype was further developed and the requirements largely incorporated, domain experts were able to suggest potential EDA tasks that could be carried out with the current prototype of MAHiCGO.

Following a period of 20-30 minutes of exploratory time with the tool in an effort to carry out tasks T1 and T2, users were asked to fill a questionnaire soliciting user opinions and reactions to the tested prototype of the visual analytic tool MAHiCGO and assessing its usefulness in user context.

IV. RESEARCH FINDINGS

The following section describes the design requirements gathered through Phases 1 and 2 followed by a snapshot of the design of the first prototype of MAHiCGO. We later present the results of the formative evaluations and describe the enhanced prototype in detail. Finally this section ends with the results of the summative evaluations including the usability themes identified and the results of the usefulness questionnaire.

A. DESIGN REQUIREMENTS

The interviews conducted during the requirement gathering phase of the co-design process, RG-I1-Mahigo/Mahi, RG-I2-Mahigo/Mahi, RG-I1-Mago, and RG-I1-Gogo (Fig. 3), resulted in the identification of a gap in the mutual representation and multi-way interaction between three genomic information: MA, Hi-C and GO.

The first interview RG-I1-Mahigo/Mahi (Fig. 3) highlighted two key concepts in understanding design requirements. The first concept is that of interactivity in the visualization of the MA plot with the ability to control the parameters for the significance level of differentially expressed genes. The second concept is that of the representation of differentially expressed genes in GO categories. The common data point combining these two requirements was the Gene ID. With a common data point, the visualization expert/ developer would be able to link the two representations in an interactive manner.

The following interview RG-I2-Mahigo/Mahi (Fig. 3) proposed the concept of linking the suggested two visualizations with Hi-C data. The visualization developer accepted the proposition because the same set of Gene ID could be used to deliver the requirement. The interview underscored the usefulness of the two-way interactivity between Hi-C and MA plots.

In addition to interactivity connecting the three visualizations, this and following interviews RG-I1-Mahigo/Mahi, RG-I2-Mahigo/Mahi, RG-I1-Mago, RG-I1-Gogo, RV-I3-Mahigo, RV-I3-Mahi, and RV-I3-Mago (Fig. 3) underscored the importance of interactivity on individual plots, with the potential to zoom, select and show more/less as part of the exploration process to select a subset of genes of interest.

We summarize the design requirements as follows:

- Simultaneous visualization and interactivity between the three visualizations - MA, Hi-C, and GO.
- Interactivity within each graph (ability to select, filter, expand/reduce, etc.).
- Ability to select and control input features/parameters and view corresponding output.
- Simple visualization of results with interpretable visual outputs/graphs.
- Little or no coding required.
- Ability to load and save data in standard format (at least csv) or images (for publication).

It is important to highlight that the users Mago and Gogo do not work with Hi-C data and therefore the incorporation of the Hi-C element as part of the GUI was not necessary for them. However, given the difficulty of finding users and the fact that MA plot and its interactive connection to GO information was a common ground between the four domain experts, we decided to incorporate Hi-C visualization as part of the prototype. At a low time and effort cost, this would have expanded the usefulness of the tool for users Mahigo and Mahi and would have added an additional layer of result validation from a different perspective as part of the analysis for Mago and Gogo.

Below is a summary of contribution of domain experts in the design features of MAHiCGO. Details and links of these contributions to the designed prototype can be found in Appendix D.

- Mahigo, Mahi, and Mago contributed to the design element of linking the MA, Hi-C and GO information in an interactive manner.
- Mahigo and Mahi contributed to the design concept of interactive MA and Hi-C plots. While Gogo was an advocate for an interactive design for the GO plot.
- Mahigo and Mahi contributed to the design concept of being able to control the parameters of visualizing gene expression information on the MA plot.
- Mago has had the major contribution in identifying the design and format of the downloaded information, such that the user can save and retrieve the selected set of genes in standard formats for further use in downstream analysis or for publications.

- Gogo introduced ShinyGO and highlighted its interactive and useful features to the co-design team.

B. PROTOTYPE DESIGN

Following the described chain of design specifications by domain experts and the verification of the gap in state-of-the-art tools by the PI, this study presents MAHiCGO as a novel tool in this specialized area of biovis. We developed MAHiCGO in Shiny from R Studio [33], compatible with Bioconductor in R, and with a GUI that requires no programming expertise.

The first version of the MAHiCGO prototype included a single tab GUI, with allocated space for the three visualizations together with a list of the selected genes. In line with user requirements, the developer designed the tool to feature gene expression results interactively on an MA plot, Hi-C results of one of the corresponding groups under study, and the related information in the current GO database. Note that this prototype at the time of the Formative Evaluations had yet not incorporated GO visualization but has allocated a space on the GUI for it.

C. RESULTS FROM FORMATIVE EVALUATIONS

Halfway through the design process, the team carried out two formative evaluation interviews with Mahigo and Mahi FE-I4-Mahigo and FE-I4-Mahi (Fig. 3). Only two domain experts were included in this phase because of time constraints.

In an effort to incorporate user feedback from Formative Evaluations, and to analyze the collected information through audio and video recordings, the co-design team listened to and transcribed results from the recordings. The highlighted usability themes included: unconventional color use for information coding for the MA plot; visualization-task mismatch, unreadable labels for the Hi-C plot, missing labels, lack of orientation and help functions, suboptimal spatial organization, and extra information presented as distraction. The developer has not yet fully completed the prototype, and therefore most of the identified themes by the users were to be incorporated as part of the work in progress.

The most critical flaw identified during these hour-long sessions was the realization that the prototype should simultaneously present the Hi-C graphs for the two groups under study in MA and not just one of them. Only when the HiC information of the two groups under study are presented is when the domain expert can carry out a proper EDA of the information.

D. ENHANCED PROTOTYPE DESIGN

In line with user requirements, the prototype has undergone further design iterations.

On the upper left corner of the first tab, the user loads the file before initiating MAHiCGO (Fig. 4, area 1). For the purposes of the evaluation exercises, the data has been preloaded. The differential expression (MA) and Hi-C datasets have been processed and uploaded to the tool as the visualization

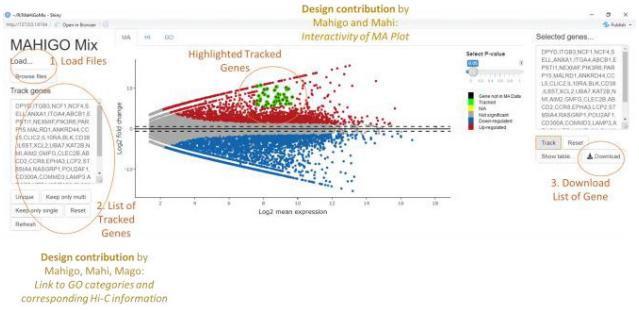

FIGURE 4. MAHiCGO GUI for MA plot highlighting the selected genes of interest in green.

expert developed the prototype and throughout the evaluation stages.

Mahigo, Mahi, and Mago contributed to the design element of linking the MA, Hi-C and GO information in an interactive manner. The interface comprises three tabs with a common pane at the leftmost part of every tab which collects and presents the set of 'tracked' genes from any of the three visualizations and which combines all three visualizations (Fig. 4, area 2). We highlighted the common set of genes in green on the three visualizations. This set will remain in the panel until the user decides to reset. The user can add gene sets to it and the tool allows for the selection of the intersection or union of the two sets, aggregating results of several selections. At any point in time, the user might want to download the selected set in a csv file (Fig. 4, area 3). Before downloading the data, the user can also view the selection in a table, with additional details about the selected genes (Fig. 5).

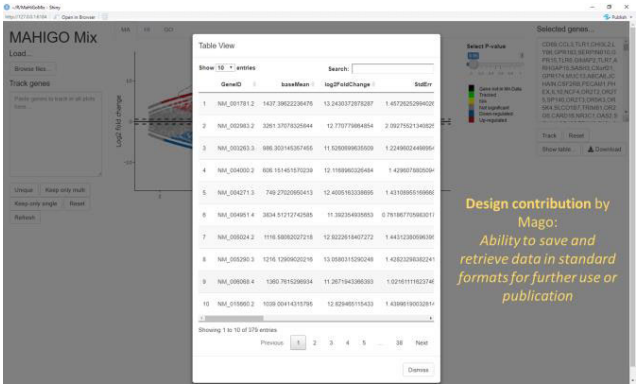

FIGURE 5. MAHiCGO allows for a table view of the selected genes ready to be downloaded for downstream analysis.

Mahigo and Mahi contributed to the design concept of interactive MA and Hi-C plots. While Gogo was an advocate for an interactive design for the GO plot. We therefore designed all three plots with interactive interfaces allowing brushing, selecting and other interactive operations. The three visualizations have been developed using ggplotly[5] which has

[5]Carson Sievert - data science (R, shiny, plotly, visualization) consultant (cpsievert.me)

the following interactive functional capabilities: download the plot as a png, zoom in/out, pan, auto-scale, lasso select, and reset axes, among others.

Appendix D presents a detailed description of the designed prototype.

E. RESULTS FROM SUMMATIVE EVALUATIONS

The team carried out four summative evaluations SE-I5-Mahigo, SE-I5-Mahi, SE-I3-Mago, and SE-I2-Gogo (Fig. 3) individually with the four participants. The evaluations lasted around an hour each and they started by guiding the users to evaluate the interface and functionalities with less focus on the data.

When we asked the domain experts to write down a specific task they would perform given MAHiCGO and its functionalities, all users were able to identify tasks that are exploratory in nature at this stage. The identified tasks presented below are largely aligned with T1 and T2.

- Mahigo: Study co-regulation of genes involved in certain pathways/biological processes. "*If you can link from GO back to differential expression that's a very interesting aspect of looking at the data; because in GO, you can have a group of genes that are supposed to be in the same functional category*".
- Mahi: What are the genes that are misrelated in Autism Spectrum Disorders? How do they contribute to the disease?
- Mago: What are the biological processes enriched among the differentially expressed genes between two conditions? - It can actually be a two-fold question starting from both sides.

"Starting from an RNA-Seq dataset with the differentially expressed genes, what is hard is to really make sense of them especially if they are a bunch of genes that you have never come across in your life. Instead of going through them one by one, especially for biological researchers who are not well-versed with bioinformatics, this is an integrated tool where with a click of a button you can go from differential expression to GO analysis."

- Gogo: What are different characteristics of basal gene expression that differentiate groups?

1) USABILITY THEMES IDENTIFIED AS USERS CARRIED OUT TASKS T1 AND T2

As highlighted in the methodology section, following identification of tasks, users were asked to spend a period of 20-30 minutes exploring the tool in an effort to carry out the two pre-defined tasks T1 and T2.

Similar to the initial open coding stage of the Grounded Theory approach, we have used these heuristics developed by Forsell and Johansson [7] and Väätäjä and colleagues [44] to categorize and present the usability topics suggested by the users during the evaluation sessions [41]. Forsell and Johansson [7] have suggested a set of 10 heuristics (H1 to H10) for information visualization as follows: spatial organization, information coding, orientation and help, data set reduction, recognition rather than recall, remove the extraneous, prompting, minimal actions, consistency, and flexibility. Väätäjä and colleagues [44] have suggested to add interaction, veracity, and aesthetics to the list [7], [44], [46], [51].

Further description of each heuristic and detailed results can be found in Appendix V. The results suggest that most comments from the Summative Evaluations were about:

- more interaction (9 comments on H11),
- flexibility (7 comments on H10),
- recognition rather than recall (6 comments on H5),
- minimal action (5 comments on H8),
- data set reduction (4 comments on H4).

The remaining comments were relating to: spatial organization (3 H1), information coding (3 H2), orientation and help (H3), prompting (3 H7), consistency (3 H9), veracity (3 H12), and aesthetics (3 H13).

We categorized the results in Appendix V by user and by tab. Points raised by domain experts with expertise relating to the corresponding visualizations will facilitate the prioritization of tasks for the next iteration of the MAHiCGO prototype design.

2) USEFULLNESS QUESTIONNAIRE RESULTS

Through a questionnaire at the end of the evaluation session (see Appendix B), evaluators asked the users if they would consider using MAHiCGO in upcoming research projects.

All users confirmed that the tool is highly useful for exploratory analysis, for the corresponding selection of a subset of genes, and for hypothesizing inferences about causal relationships of interest. Users further highlighted that in their most difficult stages of gaining an understanding of the data under study, including data integration, interactive analysis of differential expression results, analysis of Hi-C data, and correspondingly finding meaningful subsets of genes for downstream analysis, MAHiCGO would prove very useful.

Domain experts underscored that the fact that gene expression information and the other genetic information are visualized rather than listed facilitates EDA. Additionally, the interactivity within and between each and every visualization, and the fact that this can happen instantly within the same tool allowing back and forth movement between the three visualizations are very attractive features of the tool. They highlighted that these features would be helpful to both experts who are looking for exploration, and novices who are trying to have a general and simple understanding of the complex genetic information.

In summary, users have found MAHiCGO easy-to-use and novel, highlighting that they have not been aware of any current tool in the market that combines gene expression, Hi-C and GO information in a single, interactive package. User feedback included many specific comments during exploration confirming these observations. An excerpt of verbatim user comments can be found in Appendix F.

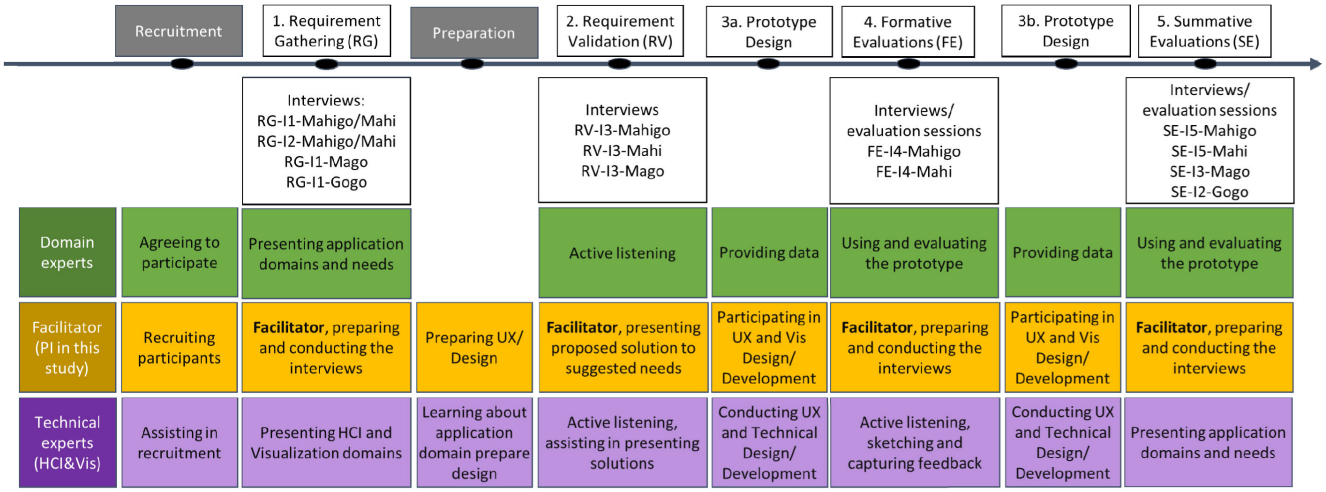

FIGURE 6. Summary of the roles of each of the domain experts, PI, and HCI and Visualization experts during the different stages of the co-design process of MAHiCGO. The role of PI as a facilitator is recommended in similar studies of co-designing computer applications of exploratory tasks in complex domains.

Mahigo and Mahi were optimistic about the usefulness of the tool conditional on improving certain features for them to trust the tool outcome and credibility for analysis purposes. Suggested improvements mainly focused on the credibility of Hi-C information, including the validity of the preprocessing code for aggregation of interaction values and missing information on parts of the chromosome. Users Mahigo and Mahi, who have been the most knowledgeable about Hi-C, suggested comparison of results from MAHiCGO with outcomes of a tool tailored by them for their research requirements and those of other state-of-the-art tools. We suggest this validation exercise as future work of MAHiCGO.

V. REFLECTIONS ON THE CO-DESIGN PROCESS

This section starts with an observation of the role of the PI in fulfilling the different needs to co-design MAHiCGO. The role is later interpreted as a facilitator. The section concludes with a reflection on the profile and generic functions of the facilitator role in co-design. This includes a broad, rather than deep, knowledge in each of HCI, visualization and application (in this case bioinformatics) domains.

A. OBSERVATION ON THE ROLE OF PI FULFILLING DIFFERENT NEEDS TO CO-DESIGN MAHICGO

The co-design team identified a large number of wide-ranging requirements at the initial stages of the requirement gathering phase. It has been challenging for the co-design team to translate the wide variety of user needs into a coherent set of requirements to be catered for by a single tool.

Where it proved challenging for the HCI researcher and the visualization expert to understand key biological concepts, the role of the PI with a broad knowledge spectrum across the three fields of biology, visualization, and HCI, proved essential in the overall success and the smooth progression of the process. The PI supported in shortlisting of the requirements by identifying the potential to connect through the Gene ID and interpreting the limitations of Shiny from R in preparation for the design.

From the start of the co-design process, the visualization expert, the HCI expert and the PI approached domain experts and raised awareness about the potential for visualization to support EDA. Nonetheless, it still proved challenging for most users to identify a visualization gap until the developer designed the first complete prototype.

Here again, we underscore the specific role of the PI in attenuating the challenges faced by domain experts to envisage the art of possible for EDA tasks.

The Requirement Gathering interviews designed by the PI started with the visualization expert introducing the wide spectrum of support interactive visualization could provide to the EDA process and demonstrated some state-of-the-art visualization tools including Shiny from R [33]. Together with the visualization expert, and through the extensive use of visualization as a means for ideation, design and evaluation, the PI prepared and conducted the interviews in a way to realize the importance of interactivity in EDA and to empower domain experts to contribute to the design.

Fig 6 provides a summary of the roles of each of the domain experts, PI, and HCI and visualization experts during the different stages of the co-design process of MAHiCGO. The roles of the PI in this research can be summarized as follows (Fig. 6):

- Connecting two complex domains and their experts through: recruiting participants; preparing and conducting interviews during Requirement Gathering and Evaluation stages
- Proposing solutions to suggested needs during Requirement Validation stages

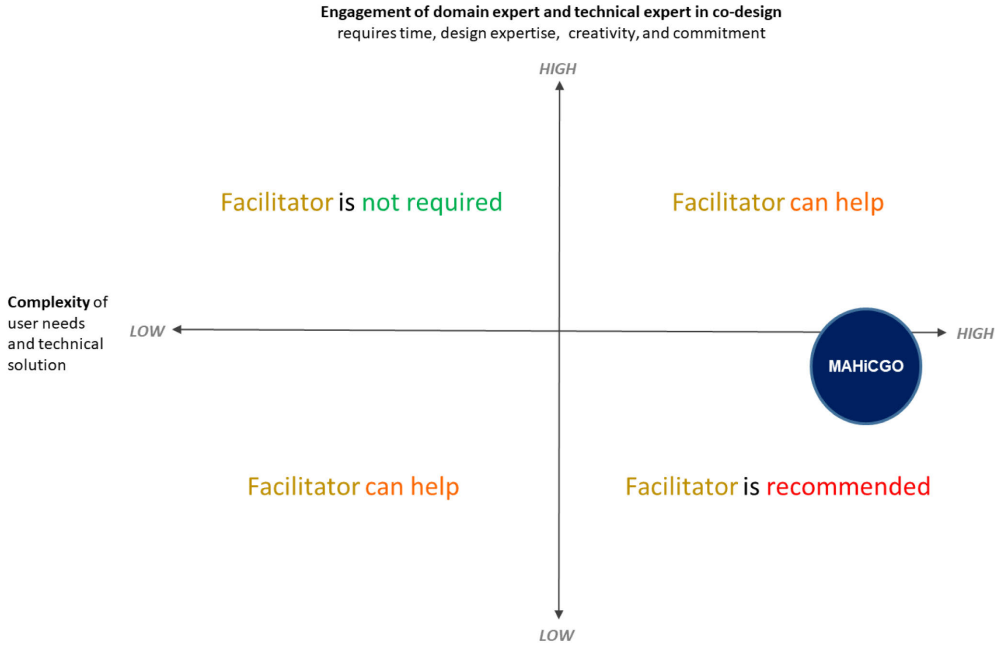

FIGURE 7. Not all design problems would need a facilitator. Sometimes application domains and the design of corresponding tasks could be complex and the experts not engaged due to lack of design knowledge or time and commitment making the co-design process less effective. We recommend the facilitator role in co-design processes of computer applications in areas of high complexity and for complex tasks. The co-design of MAHiCGO had a specific balance of a high complexity design problem and an average participation of all the domain experts thanks to the facilitator role taken by the PI.

- Participating in UX and Visualization Design/ Development.

B. PI AS A FACILITATOR

Sui [36] claims that in co-design, designers act as facilitators directing end users' level of awareness about design choices, while end-users facilitate in providing opinions and contributing with contextual experience. Despite that, in most co-design studies, team members take two key aspects for granted in identifying team structure and corresponding functional roles and responsibilities. The first is the complexity of the design problem and that of the design domain. The second is a combination of domain experts' (users) level of creativity, passion, expertise, and commitment to allocate time and attention to the co-design process (Fig. 7).

Note that in most co-design studies, designers still play a critical role in giving form to ideas [38]. There is the need for visualizing almost certainly in any area that relies on analysis of big data and there is a role for visualization experts together with the PI to play in unveiling the gaps and the needs of the domain [3].

Sometimes users or domain experts are not aware of the potential of cognitive support visualization can provide unless they are intentionally made aware of. Visualization experts could facilitate unveiling key visualization gaps essential for exploratory analysis and for supporting human cognitive abilities in exploration [3], [43].

During the interviews with the domain experts, the terms used to explain key concepts in biology was completely unfamiliar to the HCI and visualization experts and it would have otherwise required a lot of time and effort to gain the minimum necessary knowledge to keep pace with the provided information and to be able to input in the creative process of co-design.

The visualization expert stated that "basic concepts about RNASeq provided by the PI were very helpful and avoided additional discussions with domain experts. I was able to collate domain-specific questions in batches and without the need to discuss them directly with the domain experts at regularly scheduled meetings. The PI was also able to identify the underlying fundamental difference between an MA plot and HiC plot variables quicker than I could have, using more targeted search based on her basic but necessary knowledge in the application domain. Both plots were presenting gene information but in a semantically very different way and it was necessary to understand the differences to be able to link them through the gene ID."

The domain experts have reiterated the above concepts from a different perspective with Mahigo highlighting that "the PI made the design process very straightforward from our side, there was no need to go deep into the details about the biological concepts we were addressing and therefore we could concentrate on how intuitive and useful the visualization tool was. We could contribute to the design in short, well-organized and fruitful sessions." And Mahi highlighting

that ''the PI acted as the liaison between us and the Visualization/HCI experts, and helped cascade all our feedback and comments to ensure we're all aligned. She was diligent in providing us with the needed details regarding MAHiCGO's design and supported with the logistics requirements making our collaboration a very smooth and successful process.''

The HCI expert further validated the PI role as a facilitator ''addressing the knowledge and experience gap. From an HCI perspective, the first aspect realized is the complexity of this domain. There is a constant need to get back to the end-user and clarify the different concepts, terminologies, and even basic information related to RNA-seq. Being an HCI expert with more than 12 years expertise in designing for disability and almost five years' experience in designing for eHealth, it is quite clear for me now that HCI researchers would need to learn and read about the domain and at least have the minimum knowledge to be able to employ participatory design in any domain. The role of the facilitator, with the corresponding knowledge and expertise, has addressed this gap.''

The presence of the PI as a facilitator was instrumental to the fluidity of the collaborative design process. For questions from the HCI and visualization experts during remote off-meeting work, the facilitator will swiftly provide answers to questions otherwise common knowledge in the application domain without bothering the domain experts.

C. RECOMMENDATIONS ON THE NEED FOR A FACILITATOR ROLE

As both fields of biovis and design research evolve with the corresponding implications for the education of designers and researchers, we recommend including a 'facilitator' role in a co-design team for domains of high complexity and specifically for exploratory analysis. Similar to the PI in this study, the 'facilitator' will have enough knowledge in the different domains to help close the communication gap between the different experts of the design team.

As illustrated in Fig. 7, at one end of the engagement spectrum, there are domain designers with high-level of expertise, passion, creativity, and commitment to allocate time and attention to the design process; and technical (HCI/visualization) experts with wide knowledge and expertise of the specific application domain. At the other end of the spectrum, there are passive users and passive technical HCI and visualization designers.

The spectrum of the application domain and corresponding design complexity could extend from a simple design problem for a domain with low-level complexity and a widespread public knowledge and understanding, to a design for a complex task (such as EDA) in a highly complex domain with lower levels of widespread public understanding.

Not all design problems would need a facilitator. For high level of engagement of domain experts and technical experts in co-design for domains of low complexity of user needs and technical solutions, a facilitator role is not required. The facilitator role can help for both cases of high levels of domain/technical engagement for needs of high levels of complexity, and of low levels of domain/technical engagement for solution needs of low complexity. We specifically recommend the facilitator role in co-design processes of computer applications in areas of high complexity and for complex tasks such as EDA (Fig. 7).

The co-design of MAHiCGO had a specific balance of a high complexity design problem and an average participation of all the domain experts thanks to the facilitator role taken by the PI (Fig. 7). The facilitator role of the PI made the co-design process successful by filling the gaps that would have occurred otherwise by lack of time to gain the missing knowledge of the co-design team members of each other's fields.

D. THE PROFILE OF THE FACILITATOR

The field of design research is evolving [28], [38]. Sanders & Stappers highlight that as new roles shape, the future of co-design language will need to fill the gap in the functional role for communication between the design team, the various levels of 'user groups' and other stakeholders [28]. The field of bioinformatics is similarly evolving [26].

To produce optimal results, a co-design process requires from all participants a good mutual understanding of each other's field of expertise, represented in ellipses in Fig. 8.

The area of the ellipse, roughly proportional to experience, represents the total amount of knowledge acquired by the expert so far. Knowledge (the area of the ellipse) can as well be interpreted in both in breadth (angular across scientific domains) and in depth (radial along one domain axis).

As time is finite, an expert will reach his/her boundary of knowledge in his/her field (light blue dashed circle in Fig. 8) by narrowing and focusing on that specific domain earlier in career. Within the same amount of experience or time, a Facilitator (golden ellipse in Fig. 8) would have a lower depth in each domain, but a broader range of expertise across all participant domains, and therefore a larger level of mutual understanding is attained during the co-design process. Without a Facilitator, the range of knowledge reached by co-design participants (brown dashed circle in Fig. 8) would be lower and could be insufficient to solve complex problems with advanced techniques. This could lead to a failure of the co-design process for lack of engagement of the participants. Without an application domain expert (green ellipse), the problem to solve would be trivial and not relevant to the state-of-the-art of the applied domain. Without a technical expert (blue or red ellipse), the technical solution to the complex problem would not be optimal.

Therefore, HCI and visualization experts (Technical Experts), Application Domain Experts and Facilitator roles are all important for co-design to succeed when a complex problem requiring advanced solutions are expected. The facilitator role will support in the following functions:

- Connect two complex domains and their experts

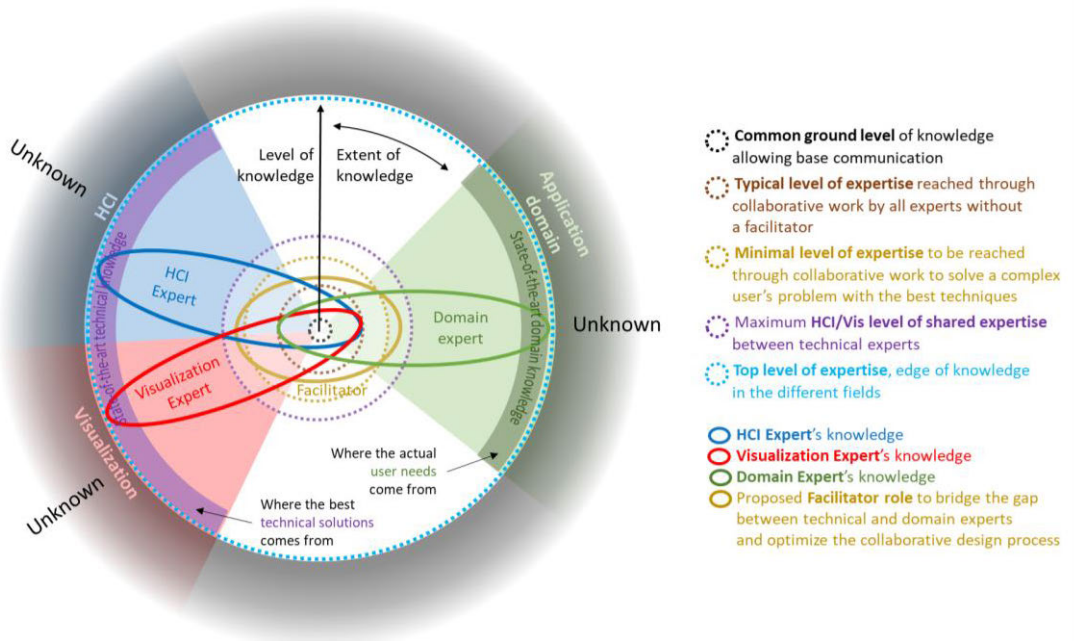

FIGURE 8. We illustrate how the Facilitator role can help bridge the knowledge gap between HCI and visualization experts (Technical Experts) and the end-users (Application domain experts) in a co-design process. To be optimal, a co-design process requires from all participants a good mutual understanding of each other's field of expertise. Each ellipse represent the knowledge both in breadth (angular across scientific domains) and in depth (radial along one domain axis). The area of the ellipse represent the total amount of knowledge acquired by the expert so far, so it is roughly proportional to the experience: a Facilitator (golden ellipse) would have a lower depth in each domain, but a broader range of expertise across all participant domains, so a larger level of mutual understanding is attained during the co-design process (golden dashed circle).

- Participate in the HCI/visualization design process informing the process with common knowledge of the application domain
- Propose solutions to the suggested needs through UX/Design development.

Competencies and the associated skill set of the Facilitator role can help bridge the knowledge gap between Technical Experts and Application domain experts in a co-design process. These include:

- Broad range of knowledge across the application domain and the HCI/ visualization domains, rather than deep technical expertise in any one of them
- Communication skills to conduct the interviews and experimental sessions, to help on the design and to maintain an active collaboration among the different co-design team members
- Empathy with both domain experts and HCI/ visualization experts throughout the co-design process.

VI. CONCLUSION

HCI research in visualization of biological data analysis, and specifically interactive visualization in exploratory data analysis, has a huge potential in supporting researchers in understanding biological processes, applying bioinformatics techniques, and exploring for new discoveries. Scalability and complexity are two challenges facing biological research. Visualization will expand the user's cognitive capabilities, and HCI research will ensure the usability of the developed tools to support the user.

As highlighted by Batch & Elmqvist [3] and others previously, we reiterate the importance of the deployment of visualization and visualization expertise to expand cognitive abilities as a means for ideation, design and evaluation.

Currently, most of the scientific explorations within the fields of RNA-Seq gene expression and the Hi-C information happens in standalone tools with little or no holistic interaction between the results of gene expression data, that of Hi-C data and potential explorations within the rich knowledge environment of the GO project.

We have described the process of developing and evaluating a prototype of MAHiCGO, a novel visualization tool to create a multi-way interactive dashboard between three visualizations. The first visualization is the MA Plot for the visualization of RNA-Seq differential gene expression between two groups; the second is the Hi-C Plot for the visualization of interactivity of genetic material in the 3D cellular space; and the third is the Gene Ontology representation of genetic data. The authors of this study have recently published a paper for an enhanced MA Plot to ease exploratory analysis of transcriptomic data [32].

As an immediate future work, it will be important to prioritize the enhancement of the current version of the tool by incorporating future enhancements highlighted as part of the evaluations of this HCI research. We summarize further suggested enhancements in Appendix G.

Engagement of domain experts from the start throughout the whole process through co-design is key. We further recommend demonstrating real usability and continued evaluation and improvement of the prototype through longitudinal studies in real setting.

In addition to designing a novel tool for the simultaneous visualization of MA plot, Hi-C plot and GO plot in an interactive manner, the main contribution of this work is the recommendation in having a 'Facilitator' as part of the co-design team. Similar to the PI in this research, the Facilitator will support in narrowing the communication gap between domain experts and technical developers to understand design requirements. The recommendation is specifically useful for the design of computer applications for complex exploratory tasks and in highly complex domains such as bioinformatics. Having a facilitator will allow to artificially increase the domain knowledge of the team members in each other's fields. This, in turn, will expand their capability to co-create and will empower team members to co-design.

ACKNOWLEDGMENT

The authors would like to thank the immense support of Dr. Borbola Mifsud, Ms. Dina Aljogol, Dr. Luis Saraiva, and Dr. Susie Huang through their expertise for the completion of this work throughout the process.

APPENDIX A
STATE-OF-THE-ART BIOVIS TOOLS

This section is a result from the interviews with domain experts followed by verifying identified gaps through literature review.

Requirement Gathering interviews have identified that as part of the conventional data analysis workflow, our users mainly resort to Galaxy and SeqMonk in the post-laboratory data preparation stages to perform functions such as Quality Control of the generated data, sequence alignment, file format conversion, filtering and visualization of regions against annotated genomes (Personal Communications, 2019-2020).

Not being satisfactory for general analyses and visualization requirements, additional stages or other tools are usually required for the following stages of exploratory analysis and the corresponding selection of a subset of genes for repeated exploration and downstream analysis [50] (Personal Communications, 2019-2020).

In Stage B, mappers are used for quantification and separate tools are generally used for gene expression analysis afterwards. Tools identified for this stage by our users include: HT-Seq, STAR, TopHat2, DESeq2, EdgeR, Cufflinks, Kallisto, Salmon, Gothic, Chicago, and Juicer. Tools such as HiCPlotter and HiGlass have been identified to visualize Hi-C data (Personal Communications, 2019-2020).

Resulting gene expression (or Hi-C) data tables are mostly explored in spreadsheet applications such as Excel, users have said. They might later be visualized in genome browsers, like UCSC and IGV, popular for visualizing data on a genome scale. Subsequent calculations are done in either spreadsheets or programming or scripting environments which require computational expertise to learn and then to keep up to speed with field advancements [50] (Personal Communications, 2019-2020).

Following and as part of the analysis in Stage B, a subset of genes is selected in Stage C for further analysis. A user has identified that gene lists from literature or from known pathways could usually be manually selected at this stage.

As identified by the users in this research, Stage D includes the most research-specific analysis. This is where the researcher needs to perform downstream analysis including hypotheses generation through different means including enrichment analysis and pathway analysis among others. For these reasons, in addition to the use of available biological ontology and visualization tools and packages - including GO, KEGG, Reactome, STRING, clusterProfiler, GSEA, Profiler - this stage might require tailor-made plots, usually in R, generated by the researcher to answer specific questions including tailored summary representation and visualization of enriched term (Personal Communications, 2019-2020).

In addition to the visualization tools and packages identified by users during interviews, we have looked at the state-of-the-art tools to visualize MA, Hi-C and GO genomic information and studied a selection of them. HiCPlotter integrates genomic data with Hi-C interaction matrices by side-by-side presentation of Hi-C interaction matrices of several experimental conditions, in addition to several tracks of different genomic features including RNA-Seq data, ChIP-Seq data and others but lacks integration with GO information [2]. HiGlass[6] is a state-of-the-art zoomable, interactive visualization of Hi-C information. HiCcompare [35], a freely available R Bioconductor package, is what we have relied to visualize Hi-C data in MAHiCGO. REVIGO and QuickGO are among the state-of-the-art tools for the visual representation of GO information; while the Augmented Exploration of the GO with Interactive Simulation, or AEGIS,[7] a recently developed 2019 open-source tool by a group of Stanford researchers, provides novel visualizations of GO terms [1], [5], [28], [37]. ShinyGO complements existing enrichment tools with graphical visualization of enrichment results without the need for programming in R and also provides an application program interface access to KEGG and STRING for the retrieval of pathway diagrams and protein-protein interaction networks [9]. We have used ShinyGO to visualize GO information in MAHiCGO.

APPENDIX B
USABILITY QUESTIONNAIRE DURING SUMMATIVE EVALUATION INTERVIEWS

To understand usefulness of MAHiCGO to the participant domain experts, the following questions were asked during the summative evaluations:

[6]http://higlass.io
[7]http://aegis.stanford.edu/

- Tell me about 2-3 interesting data analysis exercises you have conducted lately. What data did you use? How did you get the data and how did you transform it before analysis?
- Can you describe points in these processes where gaining a complete understanding of the data has been the most difficult? What makes insight most difficult at these points in the process?
- Which of the aforementioned stages do you think MAHiCGO will be useful in?
- Can you trust MAHiCGO and the statistics it generates? Do you think the data is reliable and credible?
- Does MAHiCGO support you to group genes and relationships of interest to infer causal relationships?
- Would you be considering using MAHiCGO in your research or in a scenario you would be willing to describe?

APPENDIX C
QUESTIONS ASKED DURING REQUIREMENT GATHERING INTERVIEWS

To understand user requirements, the following questions were asked during the preliminary interviews:

- "How do you carry out a typical RNA-Seq analysis? Please provide a step-by-step explanation."
- "To shape a deeper understanding of the specific type of analysis you would be willing to see in the end-product, discuss in more detail a typical downstream analysis journey that you currently follow once the differentially expressed genes are identified from an RNA sequencing. We would ideally like to see a demonstration on screen in an actual programming environment you work in."
- "Demonstrate how, when, and in what format do you discard the less-significant differentially expressed genes and carry on with the more-significantly expressed ones in the downstream analysis workflow."
- "If possible and available, provide names of softwares/tools you have encountered that provide information, and further corresponding analysis, about where in the genome the differentially expressed genes are."

APPENDIX D
DESCRIPTION OF THE ENHANCED MAHICGO PROTOTYPE DESIGN

In line with user requirements, the prototype has undergone further design iterations. We have banked our tool on existing packages:

- ggmaplot[8] for the development of the MA plot.
- HiCcompare [35] for the development of Hi-C plots.
- ShinyGO[9] for the visualization of GO information.

MAHiCGO codes gene identification information in gene names, the conventional method of referring to genes. The algorithm used in the prototype converts all corresponding information coded in Ensemble or other codes into gene names. The rest of this sub-section will present the main components of the GUI.

A. TAB 1 – MA PLOT

On the first tab, the MA plot visualizes the differential gene expressions, plotting 'log2 mean expression' versus 'log2 fold change' of all the genes in the two groups under study. The plot presents significantly differentially expressed genes in different colors. Significantly up-regulated genes are colored in red, while significantly down-regulated genes are colored in blue. Non-significantly expressed ones colored in grey. This overview presentation of all differentially expressed genes will allow users to spot genes of interest for further downstream analysis (Fig. 9).

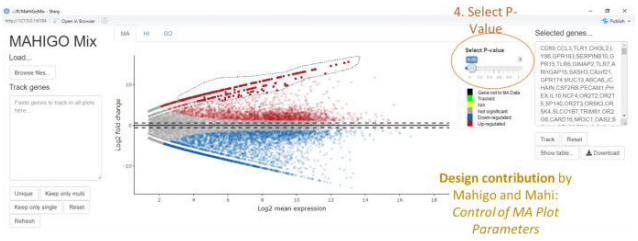

FIGURE 9. MAHiCGO allows for lasso selection of genes.

Mahigo and Mahi contributed to the design concept of being able to control the parameters of visualizing gene expression information on the MA plot. We consequently designed the MA plot such that users can select the preferred adjusted p-value for the significance level of differentiation by using the corresponding slider on the upper-right corner of the page (Fig. 9, area 4). Users will also be able to select a group of genes based on different set of cutoff criteria including: all up-regulated significantly expressed, all down-regulated significantly expressed, all above or below log-x-fold change.

Detailed information relating to the genes visualized on the MA plot can be available by hovering on the dots (genes) on the MA plot: Gene Name, Up or Down-regulated. Finally, extraneous information for the GUI, available in the dataset for back-end operations, are removed from the GUI and can be can be available on demand by downloading the data in a table format (Fig. 5). The downloaded file will contain raw information about the selected subset of genes including 'log2FoldChange', expression level 'baseMean', location on chromosome, p-value, and adjusted p-value, among others. Mago has had the major contribution in identifying the design and format of the downloaded information, such that the user can save and retrieve the selected set of genes in standard formats for further use in downstream analysis or for publications.

B. TAB 2 – HI-C PLOT

The interface on the second tab is to visualize the Hi-C information from the two groups under study. It is split

[8] ggmaplot function | R Documentation
[9] http://bioinformatics.sdstate.edu/go/

into six sections due to scalability considerations. The upper half of the page comprises three heat-map visualizations of Hi-C interactions at a chromosome-chromosome level (Fig. 10, areas a, b, and c); and the lower half correspondingly comprises three visualizations which will zoom in to the brushed/selected portions of the corresponding upper counterparts (Fig. 10, areas d, e, and f). The leftmost visualizations correspond to the Hi-C interactions within the first group under study (Fig. 10, areas a and d); while the rightmost visualizations correspond to the Hi-C interactions within the second group under study (Fig. 10, areas c and f). The middle two visualizations represent the difference in interactions of the same segments of the chromosome between the two groups (Fig. 10, areas b, e).

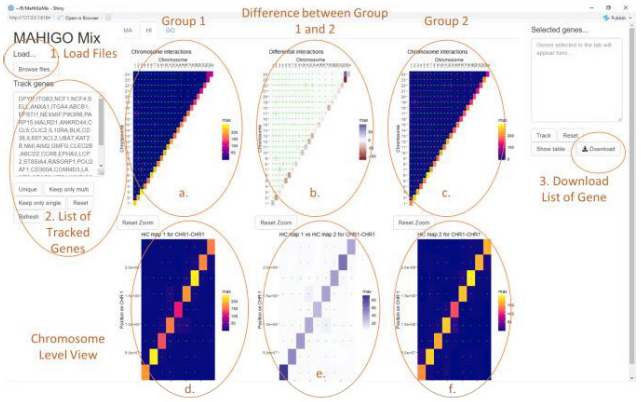

FIGURE 10. MAHiCGO GUI for Hi-C Plots. The initial chromosome-level view of the two groups under study together with the difference in Hi-C information between the two groups.

Any box in the heat-maps highlighted by a green dot identifies the existence of a tracked gene in that box. This has proven confusing for the users especially for the upper, chromosome-level graphs, where at times most chromosomes would be highlighted conveying no clear message to the user. At a zoomed level, the information will be clearer on the lower heat-maps.

The plots on the lower level have more details about the granular levels of interactions. They can be zoomed, by the order of 3, by double-clicking on a colored-box in any of the rectangles (Fig. 11). The color scale gives the intensity of the interactions. The colors are on the same scale on all six plots. Before clicking on a given box of interest, the user can get the maximum number of aggregated interactions within that region by hovering on a given box.

All lower heat-maps are zoomed at the same level; and once the zoom is at the gene level (Fig. 11, area 5), the user can brush any portion of the heat-map to select the corresponding genes. The selected genes will appear on the right hand panel and will be added to the existing list of genes if the pane was not empty (Fig. 11, area 6). The user should reset the pane if willing to start exploration from scratch. The new selection will not appear on Tab 1 or Tab 2 unless tracked. By clicking on the track button (Fig. 11, area 7), the user

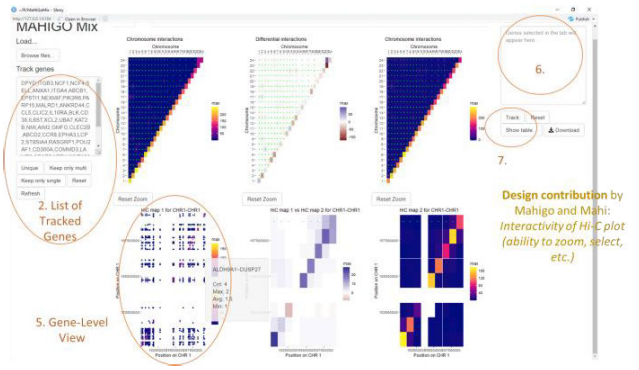

FIGURE 11. MAHiCGO GUI for Hi-C Plots – Gene-level view. By selecting a specific section of the chromosome, the user can zoom in to the selected portion of the chromosome for a more detailed and specific Hi-C information.

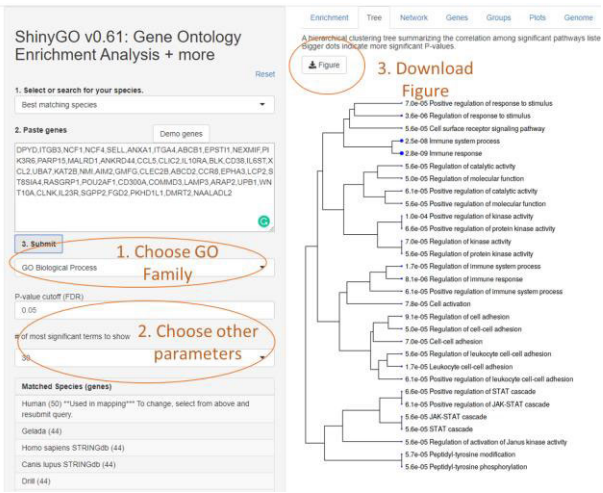

FIGURE 12. ShinyGO[10] GI for GO information–Hierarchical cluster tree view.

pushes the new set of genes to the left panel (Fig. 11, area 2). The tracked set will then appear on the MA plot or the GO visualization.

C. TAB 3 – GO PLOT

The third tab presents the selected set of genes on an interactive network visualization, allowing the user to choose the type of functional enrichment between the available options of GO Biological Process, GO Cellular Component, GO Molecular Function, KEGG, and others (Fig. 12, area 1). In addition to the functional enrichment, the user can select different options for other parameters including the p-value cutoff (FDR or False Discovery Rate), the number of most significant terms to show, edge cutoff, and others (Fig. 12, area 2).

It will be possible for users to download (Fig. 12, area 3) the different plots and tabular GO information for the selected Gene IDs including High Level GO categories (Fig. 13),

[10]http://bioinformatics.sdstate.edu/go/

TABLE 2. Usability themes for improvement highlighted through the Summative Evaluations.

Visualization	Points Raised	Participants					Heuristic Themes												
		Mahigo	Mahi	Mago	Gogo	Vis Expert	Spatial Organization	Information Coding	Orientation and Help	Dataset Reduction	Recognition rather than Recall	Remove the Extraneous	Prompting	Minimal Actions	Consistency	Flexibility	Interaction	Veracity	Aesthetics
Tab 1 – MA Plot	1. Show tracked pairs of genes from Hi-C			X							X								
	2. Allow for selection by top K expression value			X	X											X			
	3. Allow for selection by k-fold change limit			X	X					X				X			X		
	4. Allow for selection of all up-/down-regulated			X	X					X				X			X		
	5. Provide details of p, p-adj, fdr				X								X						
Tab 2 – Hi-C Plot	6. Render the heat map as square coordinates	X																	X
	7. Highlight genes in the margin to show color	X											X						
	8. Allow step by step zoom out	X		X												X	X		
	9. Include edges to the squares		X				X												X
	10. Allow filtering at k levels of interaction		X							X				X			X		
	11. Include aggregate list of genes in tooltip		X								X								
	12. Keep fixed color scales at different zoom levels		X	X											X				
	13. Remember zoom level upon return to chromosome			X										X					
	14. Show selected pairs of genes			X							X								
General GUI – All Tabs	15. Benchmark against exiting tools for accuracy of results	X	X															X	
	16. Reorder columns in downloaded tables			X			X												
	17. Assess validity of preprocessing code	X																X	
	18. Show number of genes selected				X			X			X								
	19. Include a color palette to select colors of plots			X															X
	20. Allow 'mark genes as favorite'		X					X											
	21. Include a gene search box		X										X				X		
	22. Order selected gene list alphabetically		X												X				
	23. Allow table view selection			X						X						X	X		
	24. Include guidelines for use in tooltips		X						X										
	25. Add snapshot button to save as .svg or .pdf		X											X					
	26. Do not reset plots when changing tabs		X												X	X			
	27. Allow to track a specific gene among the selected					X										X	X		
	28. Include a note area					X	X	X									X		
	29. Save meta-data of sessions					X					X								
	30. Allow history management					X					X								
	31. Allow to share sessions with remote teams					X										X	X		
	32. Display exact names of datasets above plots					X												X	

functional categories, total number and name of genes including identification of those in the selected list, enrichment FDR, network plots of the set of genes (Fig. 14) and others.

It is important to note that ShinyGO network visualizations include both static and interactive versions, allowing the user to move nodes around and hover over them to get additional information about each node (Fig. 14).

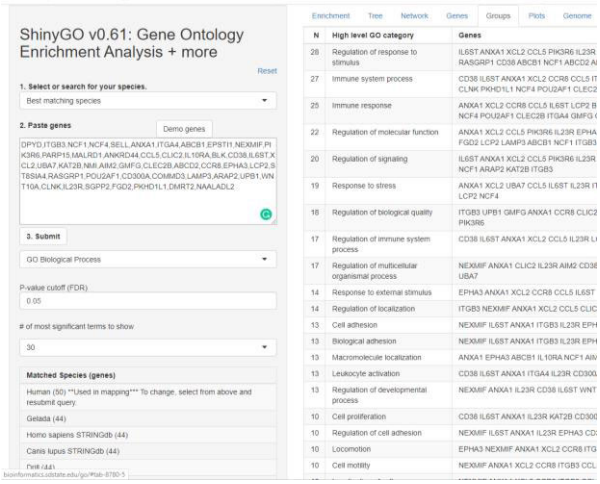

FIGURE 13. ShinyGO GUI for GO information – grouped by categories.

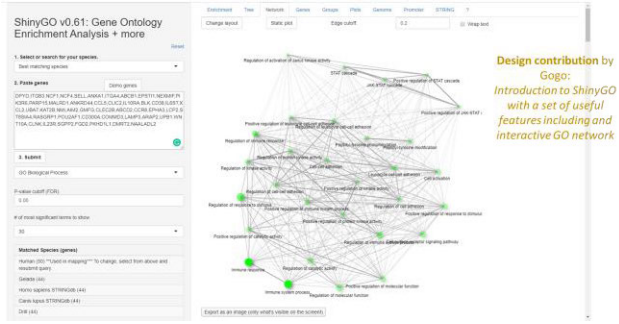

FIGURE 14. ShinyGO GUI for GO information – network plot view. ShinyGO allows for the visualization of the GO information in an interactive network plot.

APPENDIX E
USABILITY THEMES IDENTIFIED AS USERS CARRIED OUT TASKS T1 AND T2 DURING SUMMATIVE EVALUATIONS

Table 2 presents detailed usability findings highlighted by the users as they carried out tasks T1 and T2 during the final summative evaluation sessions. Note that the highlighted Points Raised are those made by participants with expertise relating to the corresponding visualizations. This will facilitate prioritization of tasks by the developer of the next iteration prototype of MAHiCGO. We have used Forsell and Johansson's [7] set of 10 heuristics in addition to the three later added by Väätäjä and colleagues [44] to categorize and present the usability themes suggested by the users during the evaluation sessions [41].

APPENDIX F
USEFULLNESS QUESTIONNAIRE RESULTS FROM SUMMATIVE EVALUATION SESSIONS

See Table 3.

APPENDIX G
FUTURE ENHANCEMENTS OF MAHICGO

Future enhancements highlighted as part of the evaluations of this HCI research can be summarized as follows:

TABLE 3. Usefulness of MAHiCGO assessed by participants.

- ***Mahigo***: "It is "nice" that we can download the detailed table on demand through the Show Table feature."
- "It looks actually very good that you were able to connect to the GO data."
- "Actually it's very nice because the set of genes were showing losing interactions on Hi-C" and the information can be confirmed on MA – "the data are aligned and make sense."
- "If you can link from GO back to differential expression that's a very interesting aspect of looking at the data; because in GO, you can have a group of genes that are supposed to be in the same functional category."

- ***Mahi***: "Even without looking at the genes, the differential interactions graph is useful and is showing some interesting interactions."
- "All the genes I thought were interesting turned out to be involved in the process of cell death - it's definitely interesting - the fact that they all have the same function is very useful."

- ***Mago***: "If I focus on the genes that are part of the signaling pathway X - we go to GO… you can always do this - What people usually do, they go to the literature, which will take a long time, come back and look for the genes one by one… things are so quick in here."
- "The two-way interactions between these three visualizations would be interesting and useful at any given phase of the analysis."
- "If a student enrolls in a master's program, just to learn bioinformatics you're talking about a year. So by the time the student learns bioinformatics and the student has no background – that's it. Here, in this tool: here is the data, go! You can't mess it up, because you're just playing – it's easy to explain, it's easy to understand. It's just great!"

- ***Gogo***: "ShinyGO is nice because the visualizations there are themselves interactive."
- "Visualization is more important than just a list of genes from enrichment analysis - it's good that shinyGO has nice representations - other tools with very useful visualizations exist as well, I would rely on those tools."
- "It's really nice, as they can individually, on a short period of time, get information about a large number of genes by just hovering over them."
- "This will be a very nice tool for someone who wants to understand a bunch of numbers that a bioinformatician has provided."
- "It's really nice that you have all these three options and that the user does not necessarily need to use them but has the options to use if needed."

Personal Communications, 2019-2020 – information gathered through interviews with users

- enhancing the algorithm and filling scalability gaps by identifying a better algorithm for loading the datasets
- improved aesthetics
- visualizing both gene and non-gene interactions on the Hi-C plot
- providing different algorithms to annotate genes and recommending the best including other pathway analysis repositories such as Reactomes.

REFERENCES

[1] (2019). *AEGIS Documentation*. [Online]. Available: http://aegis.stanford.edu/

[2] K. C. Akdemir and L. Chin, "HiCPlotter integrates genomic data with interaction matrices," *Genome Biol.*, vol. 16, no. 1, p. 198, 2015. Accessed: Apr. 9, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4576377/, doi: 10.1186/s13059-015-0767-1.

[3] A. Batch and N. Elmqvist, "The interactive visualization gap in initial exploratory data analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 278–287, Jan. 2018, doi: 10.1109/TVCG.2017.2743990.

[4] D. Christopoulos, "Peer esteem snowballing: A methodology for expert surveys," in *Proc. Eurostat Conf. New Techn. Technol. Statist.* Brussels, Belgium: Eurostat, Feb. 2009, pp. 171–179. Accessed: Apr. 9, 2020. [Online]. Available: https://www.researchgate.net/publication/255595189

[5] D. Ch and N. Skunca, "The gene ontology handbook," in *Methods in Molecular Biology*, vol. 1446, 2017th ed. Nov. 2016, pp. 3–37. Accessed: Apr. 9, 2020. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2F978-1-4939-3743-1.pdf, doi: 10.1007/978-1-4939-3743-1.

[6] A. Dix, J. Finlay, G. D. Abowd, and B. Russell, Eds., *Human-Computer Interaction*. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.

[7] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA, USA: MIT Press, 1993.

[8] C. Forsell and J. Johansson, "An heuristic set for evaluation in information visualization," in *Proc. Int. Conf. Adv. Vis. Interfaces*, New York, NY, USA, 2010, pp. 199–206. Accessed: Apr. 9, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/1842993.1843029, doi: 10.1145/1842993.1843029.

[9] S. Ge, D. Jung, and R. Yao, "ShinyGO: A graphical gene-set enrichment tool for animals and plants," *Bioinformatics*, vol. 36, no. 8, pp. 2628–2629, 2020. Accessed: Apr. 9, 2020, doi: 10.1093/bioinformatics/btz931

[10] Rudjer Boskovic Institute, Croatia. (Mar. 2017). *REVIGO*. Accessed: Apr. 9, 2020. [Online]. Available: http://revigo.irb.hr/

[11] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Moller, "A systematic review on the practice of evaluating visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2818–2827, Dec. 2013. Accessed: Apr. 9, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/6634108, doi: 10.1109/TVCG.2013.126.

[12] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer, "Enterprise data analysis and visualization: An interview study," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2917–2926, Dec. 2012. Accessed: Apr. 9, 2020. [Online]. Available: https://dl.acm.org/doi/10.1109/TVCG.2012.219, doi: 10.1109/TVCG.2012.219.

[13] A. M. Kanstrup, A. Bygholm, N. Ch, and P. Bertelsen, "Participatory design & health information technology," *Studies in Health Technology and Informatics*. Amsterdam, The Netherlands: IOS Press, 2017.

[14] P. Kerpedjiev, N. Abdennur, F. Lekschas, M. Ch, K. Dinkla, H. Strobelt, J. M. Luber, S. B. Ouellette, A. Azhir, N. Kumar, J. Hwang, S. Lee, B. H. Alver, H. Pfister, L. A. Mirny, P. J. Park, and N. Gehlenborg, "HiGlass: Web-based visual comparison and exploration of genome interaction maps," *Genome Biol.*, vol. 19, no. 1, p. 125, 2018. Accessed: Apr. 9, 2020. [Online]. Available: http://higlass.io/about, doi: 10.1186/s13059-018-1486-1.

[15] A. Kerren, K. Kucher, Y.-F. Li, and F. Schreiber, "BioVis explorer: A visual guide for biological data visualization techniques," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0187341, doi: 10.1371/journal.pone.0187341.

[16] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 9, pp. 1520–1536, Dec. 2012. Accessed: Apr. 9, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/6095544

[17] E. Lander *et al.*, "Initial sequencing and analysis of the human genome," *Int. Hum. Genome Sequencing Consortium. Nature*, vol. 409, no. 6822, pp. 860–921, 2001. Accessed: Apr. 9, 2020, doi: 10.1038/35057062.

[18] J. Lazar, H. F. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2017.

[19] C. A. J. Lazarin and L. D. A. Almeaida, "Distributed participatory design web-based groupware: Gathering requirements through braindraw," in *Proc. 15th Brazilian Symp. Hum. Factors Comput. Syst. (ICH)*, New York, NY, USA, vol. 3, 2016, pp. 1–10, doi: 10.1145/3033701.3033704.

[20] Y. Luo, P. Liu, and E. K. Choe, "Co-designing food trackers with dietitians: Identifying design opportunities for food tracker customization," in *Proc. CHI Conf. Hum. Factors Comput. Syst. Process. (CHI)*, Glasgow, U.K. May 2019, Art. no. 4. Accessed: Nov. 12, 2020, doi: 10.1145/3290605.3300822.

[21] MA Plots. (2019). *ScienceDirect a Registered Trademark of Elsevier B.V.* Accessed: Apr. 9, 2020. [Online]. Available: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/ma-plot

[22] B. Mirel, F. Eichinger, B. Keller, and M. Kretzler, "A cognitive task analysis of a visual analytic workflow: Exploring molecular interaction networks in systems biology," *J. Biomed. Discovery Collaboration*, vol. 6, pp. 1–33, 2011. Accessed: Apr. 9, 2020. [Online]. Available: https://doi.org/10.5210/disco.v6i0.3410, doi: 10.5210/disco.v6i0.3410.

[23] C. A. Ntuen, E. H. Park, and K. Gwang-Myung, "Designing an information visualization tool for sensemaking," *Int. J. Hum.-Comput. Interact.*, vol. 26, nos. 2–3, pp. 189–205, Mar. 2010. Accessed: Apr. 9, 2020, doi: 10.1080/10447310903498825. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/10447310903498825

[24] S. Nusrat, T. Harbig, and N. Gehlenborg, "Tasks, techniques, and tools for genomic data visualization," *Comput. Graphics Forum*, vol. 38, no. 3, pp. 781–805, 2019. Accessed: Apr. 9, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/cgf.13727, doi: 10.1111/cgf.13727.

[25] S. O'Donoghue, A. Gavin, N. Gehlenborg, D. Goodsell, J. Hériché, and C. Nielsen, "Visualizing biological data—Now and in the future," *Nature Methods*, vol. 7, pp. S2–S4, Mar. 2010. Accessed: Apr. 9, 2020. [Online]. Available: https://www.nature.com/articles/nmeth.f.301, doi: 10.1038/nmeth.f.301.

[26] C. Ouzounis, "Rise and demise of bioinformatics? Promise and progress," *PLoS Comput. Biol.*, vol. 8, no. 4, 2012, Art. no. E1002487. Accessed: Apr. 9, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3343106/, doi: 10.1371/journal.pcbi.1002487.

[27] V. L. Patel, T. G. Kannampallil, and D. R. Kaufman, *Cognitive Informatics for Biomedicine: Human Computer Interaction in Healthcare* (Health Informatics), 1st ed. Amsterdam, The Netherlands: IOS Press, 2015.

[28] E. B.-N. Sanders and P. J. Stappers, "Co-creation and the new landscapes of design," *Co-Des.*, vol. 4, no. 1, pp. 5–18, 2008. Accessed: Nov. 12, 2020, doi: 10.1080/15710880701875068.

[29] Rudjer Boskovic Institute, Croatia. (Mar. 2017). *REVIGO*. Accessed: Apr. 9, 2020. [Online]. Available: http://revigo.irb.hr/

[30] P. Saraiya, C. North, V. Lam, and K. A. Duca, "An insight-based longitudinal study of visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 6, pp. 1511–1522, Nov. 2006. Accessed: Apr. 9, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/17073373, doi: 10.1109/TVCG.2006.85.

[31] P. Saraiya, N. Ch, and K. Duca, "An evaluation of microarray visualization tools for biological insight," in *Proc. IEEE Symp. Inf. Vis. (INFOVIS)*, Oct. 2004, pp. 1–8. Accessed: Apr. 9, 2020. [Online]. Available: https://dl.acm.org/doi/10.5555/1038262.1038775

[32] A. Sheharyar, T. B. Yacoubian, D. Aljogol, B. Mifsud, T. D. Al, and A. M. , "An enhanced MA plot with R-Shiny to ease exploratory analysis of transcriptomic data," presented at the BioVis Redesign Challenge IEEE VIS, 2020. Accessed: Dec. 20, 2020. [Online]. Available: http://biovis.net/2020/program_ieee/ and https://arxiv.org/abs/2012.04411

[33] (2017). *Shiny From R Studio*. Accessed: Apr. 9, 2020. [Online]. Available: https://shiny.rstudio.com/

[34] A. Slingsby and J. Dykes, "Experiences in involving analysts in visualisation design," in *Proc. BELIV Workshop Beyond Time Errors-Novel Eval. Methods Vis. (BELIV)*, New York, NY, USA, 2012, pp. 1–4. Accessed: Apr. 9, 2020, doi: 10.1145/2442576.2442577.

[35] C. S. Stansfield, K. G. Cresswell, V. I. Vladimirov, and M. G. Dozmorov, "HiCcompare: An R-package for joint normalization and comparison of HI-C datasets," *BMC Bioinf.*, vol. 19, no. 1, p. 279, 2018. Accessed: Apr. 9, 2020. [Online]. Available: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2288-x, doi: 10.1186/s12859-018-2288-x.

[36] K. W. M. Sui, "Users' creative responses and designers' roles," *Des. Issues*, vol. 19, no. 2, pp. 64–74, 2003. Accessed: Jan. 3, 2021, doi: 10.1162/074793603765201424.

[37] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc, "REVIGO summarizes and visualizes long lists of gene ontology terms," *PLoS ONE*, vol. 6, no. 7, Jul. 2011, Art. no. e21800.

[38] S. Taffe, "The hybrid designer/end-user: Revealing paradoxes in co-design," *Design Stud.*, vol. 40, pp. 39–59, Sep. 2015. Accessed: Jan. 18, 2020, doi: 10.1016/j.destud.2015.06.003.

[39] (Jul. 2019). *The Gene Ontology Resource*. Accessed: Apr. 9, 2020. [Online]. Available: http://geneontology.org/

[40] A. Thomer, M. Twidale, J. Guo, and M. Yoder, "Co-designing scientific software: Hackathons for participatory interface design," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst. (CHI EA)*, New York, NY, USA, 2016, pp. 3219–3226, Accessed: Apr. 9, 2020, doi: 10.1145/2851581.2892549.

[41] Y. T. Ch, M. Birks, and K. Francis, "Grounded theory research: A design framework for novice researchers," *SAGE Open Med.*, vol. 7, Jan. 2019, Art. no. 2050312118822927. Accessed: Apr. 9, 2020, doi: 10.1177/2050312118822927.

[42] C. Tominski, C. Forsell, and J. Johansson, "Interaction support for visual comparison inspired by natural behavior," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2719–2728, Dec. 2012. Accessed: Apr. 9, 2020. [Online]. Available: http://web.cse.ohio-state.edu/~machiraju.1/teaching/CSE5544/Visweek2012/infovis/papers/tominski2.pdf, doi: 10.1109/TVCG.2012.237.

[43] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA, USA: Addison-Wesley, 1977.

[44] H. Väätäjä, J. Varsaluoma, T. Heimonen, K. Tiitinen, J. Hakulinen, M. Turunen, H. Nieminen, and P. Ihantola, "Information visualization heuristics in practical expert evaluation," in *Proc. Beyond Time Errors Novel Eval. Methods Vis.*, New York, NY, USA, 2016, pp. 36–43. Accessed: Apr. 9, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/2993901.2993918, doi: 10.1145/2993901.2993918.

[45] F. S. Visser, P. J. Stappers, R. Van der Lugt, and E. B.-N. Sanders, "Contextmapping: Experiences from practice," *CoDesign*, vol. 1, no. 2, pp. 119–149, 2005. Accessed: Nov. 12, 2020, doi: 10.1080/15710880500135987.

[46] R. Williams, J. Scholtz, L. M. Blaha, L. Franklin, and Z. Huan, "Evaluation of visualization heuristics," in *Proc. Int. Conf. Hum.-Comput. Interact., HCI Hum.-Comput. Interact. Theories, Methods, Hum. Issues*, in Lecture Notes in Computer Science, vol. 10901, 2018, pp. 208–224. Accessed: Apr. 9, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-91238-7_18

[47] Y. Yao, J. R. Basdeo, S. Kaushik, and Y. Wang, "Defending my castle: A co-design study of privacy mechanisms for smart homes," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, Glasgow, U.K., May 2019, p. 12. Accessed: Nov. 12, 2020, doi: 10.1145/3290605.3300428.

[48] S. Wilson and R. Maclean, *Research Methods and Data Analysis for Psychology*. New York, NY, USA: McGraw-Hill, 2011.

[49] J. S. Yi, Y. Kang, and J. Stasko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1224–1231, Nov. 2007, Accessed: Apr. 9, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/4376144, doi: 10.1109/TVCG.2007.70515.

[50] H. Younesy, T. Möller, M. C. Lorincz, M. M. Karimi, and S. J. Jones, "VisRseq: R-based visual framework for analysis of sequencing data," *BMC Bioinf.*, vol. 16, Aug. 2015, Art. no. S2. Accessed: Apr. 9, 2020, doi: 10.1186/1471-2105-16-S11-S2.

[51] T. Zuk, L. Schlesier, P. Neumann, M. Hancock, and S. Carpendale, "Heuristics for information visualization evaluation," in *Proc. AVI Workshop Beyond Time Errors Novel Eval. Methods Inf. Vis.*, 2006, pp. 1–6. Accessed: Apr. 9, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/1168149.1168162, doi: 10.1145/1168149.1168162.

TALAR B. YACOUBIAN was born in Beirut, Lebanon. She received the B.S. degree in mathematics from Haigazian University, Beirut, in 2008, the M.S. degree in financial economics from the American University of Beirut, in 2010, and the M.D.A. degree in health management from Hamad Bin Khalifa University, Doha, in 2020.

In 2010, she was a Research Assistant with the Department of Economics, American University of Beirut. Her research interests include cross-domain collaboration and using visual analytics to support solving complex problems.

DENA A. AL-THANI received the M.Sc. degree (Hons.) in software engineering from the University of London, in 2009, and the Ph.D. degree in computer science in 2016, her thesis was titled Understanding and Supporting Cross-Modal Information Seeking. Her thesis in human computer interaction (HCI) investigated visually impaired and sighted people's collaborative computer use and proposed technical approaches to support it. Following her graduation, she managed the online portal and integration platforms for Ooredoo-Qatar before joining Queen Mary University of London's Ph.D. studentship program to continue research in her field of interest. The thesis explored the under-investigated area of cross-modal interaction and inclusive design and evaluation. As an HCI Researcher, she has aims to explore how we can better design technologies to support the daily lives of people with disabilities. In addition to her research work at the Queen Mary University of London, she received a postgraduate certificate of Learning and Teaching in Higher Education. She is currently an Assistant Professor with the College of Science and Engineering, Hamad Bin Khalifa University. She is also an Associate Member of the Higher Education Academy in the U.K.

MICHAËL AUPETIT received the Ph.D. degree in industrial engineering from the Grenoble National Polytechnic Institute (INPG), in 2001, and the Habilitation for Research Supervision (HDR) degree in computer science from Paris 11 Orsay University, in 2012.

He has been a Senior Scientist with the Qatar Computing Research Institute, since 2014. He works with the Social Computing group. Before joining QCRI, he worked for ten years as a research scientist and a senior expert of data mining and visual analytics at CEA LIST in Paris Saclay, where he designed decision support systems to solve complex industrial problems in health and security domains. His research interests include the use and usability of machine learning, topological inference, and information visualization techniques together under the visual analytics umbrella, to bridge the gap between machines and humans for exploring multidimensional data. He uses his expertise to address top priority challenges of Qatar in eHealth. He initiated and co-organized the first VisCHI conference in Qatar, and five workshops in international conferences. He is a PC Member of IEEE VAST conference, and has reviewed 100 articles for top journals and conferences, has more than 70 publications, and holds two WO and one EP patents.