Exploring and visualizing multidimensional data in translational research platforms

William Dunn, Jr, Anita Burgun, Marie-Odile Krebs and Bastien Rance

Corresponding author: Bastien Rance, Biomedical Informatics and Public Health Department, University Hospital Georges Pompidou, AP-HP, Paris, France and INSERM, Centre de Recherche des Cordeliers, team 22: Information Sciences to support Personalized Medicine, Université Paris Descartes, Sorbonne Paris Cité, Faculté de médecine, Paris, France. Tel.: +33 1 56 09 59 85; Fax: +33 1 56 09 20 52; E-mail: bastien.rance@egp.aphp.fr

Abstract
The unprecedented advances in technology and scientific research over the past few years have provided the scientific community with new and more complex forms of data. Large data sets collected from single groups or cross-institution consortiums containing hundreds of omic and clinical variables corresponding to thousands of patients are becoming increasingly commonplace in the research setting. Before any core analyses are performed, visualization often plays a key role in the initial phases of research, especially for projects where no initial hypotheses are dominant. Proper visualization of data at a high level facilitates researcher’s abilities to find trends, identify outliers and perform quality checks. In addition, research has uncovered the important role of visualization in data analysis and its implied benefits facilitating our understanding of disease and ultimately improving patient care. In this work, we present a review of the current landscape of existing tools designed to facilitate the visualization of multidimensional data in translational research platforms. Specifically, we reviewed the biomedical literature for translational platforms allowing the visualization and exploration of clinical and omics data, and identified 11 platforms: cBioPortal, interactive genomics patient stratification explorer, Igloo-Plot, The Georgetown Database of Cancer Plus, tranSMART, an unnamed data-cube-based model supporting heterogeneous data, Papilio, Caleydo Domino, Qlucore Omics, Oracle Health Sciences Translational Research Center and OmicsOffice VR powered by TIBCO Spotfire. In a health sector continuously witnessing an increase in data from multifarious sources, visualization tools used to better grasp these data will grow in their importance, and we believe our work will be useful in guiding investigators in similar situations.

Key words: high-dimensional data; omics; data analytics; visualization; translational research

Introduction

Background
The continued digitization of our world along with recent advances in technology are providing researchers with data at an unprecedented rate in a variety of fields such as molecular biology, business and government [1]. Big data in general is typically challenged by five Vs (sheer volume, velocity data are received and sent, variety of formats and types, questions of veracity and ability to turn raw data into valuable information), and medical research data are no exception. The technological
advances that have followed in the wake of the next-generation sequencing (NGS) experiments at the turn of the 21st century [2] have given rise to the production of ‘big-data’ at a scale never seen before. As a result of this recent abundance of data, some have proposed that fundamental paradigms in a variety of domains—especially molecular biology—have shifted to data-driven analysis and visualization leveraging computational power and computer science [3, 4].

Growing need for multidimensional visualizations in health research

In a research environment focused increasingly on high throughput, a common challenge is the comprehensive visualization of data, an important step for any extensive exploration of the data. In Heer et al. [1], apart from providing a thorough review of emerging visualization techniques for big data, the authors outlined several benefits of quality visualization such as facilitating our ability to see patterns, trends and outliers, improving comprehension, memory, and decision-making and finally adding aesthetic appeal to engage a wider audience in data exploration and analysis.

In health care or clinical research settings, visual analytics is especially useful in studying parameters across patients when no clear hypotheses are immediately available [5]. Whereas traditional analysis of heterogeneous or multidimensional cohort data with partial overlap usually involves limiting attention to certain subsets (inevitably leading to loss of the overall sense of relationships between different modalities), a thorough visualization can provide a more complete picture, ultimately allowing a more comprehensive study of the data that improves hypothesis and research workflow [6]. As a result, systematic organization of research data can facilitate translational science and jump-start drug discovery [7], contribute to patient stratification and personalized medicine [8] and ultimately improve quality health care [9].

Driving motivation for the review

Quality visualization can be applied to any of the numerous domains where big data has recently affected the health-care arena such as, among others, managing cost, improving quality improvement, monitoring patients for clinical deterioration and improving treatment efficiency in emergency care [10–13]. In clinical research, multidimensional data can be used to help segment patients or elucidate disease pathway. This has most notably been seen in oncology with large data sets containing various genomics and clinical data for thousands of cancer patients such as The Cancer Genome Atlas (TCGA [14]) or the International Cancer Genome Consortium ([15]). However, multi-omics research has extended into a wide variety of fields such as dementia and Alzheimer’s disease (Alzheimer’s Disease Neuroimaging Initiative [16]), autism spectrum disorder (National Database for Autism Research [17]), psychiatric diseases (Psychiatric Genomics Consortium [18]), as well as for rare diseases (RD-Connect [19]). To better explore and take advantage of these rich, diverse data sets, a comprehensive exploration of data using efficient visualization that allows experts to seamlessly explore heterogeneous data on demand is required.

Multidimensional visualization basics

While basic statistics visualizations such as histograms, bar charts, line graphs or scatter plots typically suffice for one- or two-dimensional data, complex multidimensional data pose more challenges to researchers. The central question is usually how to better grasp the rich multivariable data and their relations contained in data sets with hundreds or thousands of patients or variables.

A variety of techniques ranging from simple box plots to complex radial tree layout diagrams [20] exist to better visualize multiple variables of a multidimensional data set. We have provided a brief sampling of these techniques based on several variables from a local study in Figure 1. For example, interactive, filterable, dynamic pivot tables can allow for a variety of visualizations for multidimensional data. Correlation matrices using multiple scatter plots show an additional insight into the interaction between variables. In addition, heatmaps are commonly used for multidimensional data, especially in genetic research with expression, pathway or molecular abundance data and involve a matrix where each cell is colored according to a gradient and is often clustered by samples [22]. Heatmaps and other visualizations are available in a wide variety of software such as R, Matlabb, SAS®, as well as to users without programming knowledge through programs with intuitive user interfaces (e.g. ClustVis [23], HemI [24]).

Another increasingly common technique for visualizing the relationships between variables in multidimensional data sets is parallel coordinates. Here, vertical axes corresponding to each variable scaled to a common height are placed next to each other and connected with lines representing different samples [25]. This technique has been enhanced by tools such as scatter plot matrix overlay [26], proximity-based shading [27] and clustering methods that eliminate overplotting [28]. One particular application of parallel coordinate visualization in current research is Dynamics Visualization based on Parallel Coordinates, which uses multidimensional methods to visualize complex and dynamic biochemical networks to better understand disease mechanism and ultimately to derive effective treatment strategies [29].

In many cases, multidimensional visualizations can be combined with each other. For example, visualizations can be constructed to provide elegant high-level representations of large multi-omics studies containing billions of data points arising from multiple genetic experiments and clinical and demographic data from hundreds of patients [30–32]. For instance, OmicCircos [33] is an R package that produces circular plots capable of integrating expression, copy number variations (CNV) and protein fusions as well as visualizations of statistics that compare data across these sources. This allows researchers a high-level view that may facilitate the understanding of complex diseases such as cancer or psychiatric diseases. Two other interesting R packages that integrate multi-omics with visualizations are coMET [34], which incorporates epigenetic results and other types of genomic data such as expression profiles, and caOmicsV [35], which also provides several options of viewing various genomic data side-by-side other phenotypic data.

The field of data visualization is immense. Dedicated tools and libraries have been developed and exist through a rising number of open-source and fee-based platforms. For example, many scientists rely on various programming languages or statistics packages with open-source data-visualization capabilities such as R [36] or Python. Matplotlib [37]. More and more researchers are turning to JavaScript graphics libraries to enhance visualization with dynamic capabilities. Such libraries include Highcharts [38], Chart.js [39], D3.js [40], JavaScript InfoVis Toolkit [41] and D3.js (Data-Driven Documents [42]) (for comprehensive overview and side-by-side comparison of these libraries see [43]). In sum, impressive techniques have been developed to answer to the clear need for strong data visualization in health-care research.

However, such tools and techniques are not easily accessible to the clinician or biologist end users. R packages or Python
library are easy to leverage for a bioinformatician, but the knowl-
dge gap is often too wide for biologists and clinicians without
a background in bioinformatics or biostatistics. A common chal-
lenge is finding these visualizations seamlessly incorporated
within a translational research platform without the need for
complicated backend programming. Such systems would open
the door to all members of the clinical research team, not only
those with programming backgrounds, a common theme in
contemporary translational bioinformatics [44].

In this work, we will review the tools available to researchers
and clinicians that fill this gap and provide intuitive visualization
solutions for multidimensional medical research data with easy-to-use graphical user
interfaces. Therefore, we have strived to include a wide variety
of tools with slightly different dedicated domains, structure and
capabilities and availabilities. The first three platforms identified
that respected these inclusion criteria were three platforms
from the previous review [cBioPortal, The Georgetown Database
of Cancer (G-DOC) Plus and tranSMART]. We then searched sci-
entific literature available through PubMed [46] using Medical
Subject Heading terms and free-text search, and subsequently
identified 367 articles potentially describing visualization for
heterogeneous data [45]. In the current project, we searched for
systems (i) that accept a variety of data types (and at least clin-
ical and omics data), (ii) that feature data visualization function-
alities and (iii) that provide researchers with data analysis or
statistical functionalities. We are interested in characterizing a
comprehensive current landscape of tools that can be used in
translational research to provide visualizations for multidimen-
sional medical research data with easy-to-use graphical user
interfaces. Therefore, we have strived to include a wide variety
of tools with slightly different dedicated domains, structure and
capacities and availabilities. The first three platforms identified
that respected these inclusion criteria were three platforms
from the previous review [cBioPortal, The Georgetown Database
of Cancer (G-DOC) Plus and tranSMART]. We then searched sci-
entific literature available through PubMed [46] using Medical
Subject Headings terms and free-text search, and subsequently
identified 367 articles potentially describing visualization for
heterogeneous data (PubMed queries and literature search, de-
tails are available in Supplementary Table S1). We identified
three new platforms through this step, and one from citations

Materials and methods

Literature review methods

Our literature review can be seen as a follow-up to our previous
article reviewing translational research platforms integrating
heterogeneous data [45]. In the current project, we searched for
systems (i) that accept a variety of data types (and at least clin-
ical and omics data), (ii) that feature data visualization function-
alities and (iii) that provide researchers with data analysis or
statistical functionalities. We are interested in characterizing a
comprehensive current landscape of tools that can be used in
translational research to provide visualizations for multidimen-
sional medical research data with easy-to-use graphical user
interfaces. Therefore, we have strived to include a wide variety
of tools with slightly different dedicated domains, structure and
capacities and availabilities. The first three platforms identified
that respected these inclusion criteria were three platforms
from the previous review [cBioPortal, The Georgetown Database
of Cancer (G-DOC) Plus and tranSMART]. We then searched sci-
entific literature available through PubMed [46] using Medical
Subject Headings terms and free-text search, and subsequently
identified 367 articles potentially describing visualization for
heterogeneous data (PubMed queries and literature search, de-
tails are available in Supplementary Table S1). We identified
three new platforms through this step, and one from citations

Figure 1. A sampling of commonly used visualization techniques for multidimensional data using a subset of data in our data set compiling data from three groups of
patients Var1, Var2 and Var3 are neurocognitive dimensions, Var4 and Var5 are psychopathological dimensions and Var6 is a global genetic index. Specific visualiza-
tions used are (A) dynamic pivot table (using R ‘pivotTable’ package), (B) correlation matrix (using R ‘PerformanceAnalytics’ package), (C) Heatmap clustered by rows
and columns (using R ‘gplots’ package), (D) 3D scatterplot using color and size (using R ‘scatterplot3d’ package) and (E) parallel coordinates showing all data (using d3
Javascript library ‘d3.parcoords.js’ [21]). A colour version of this figure is available at BIB online: https://academic.oup.com/bib.
results for one of the corresponding publications. To completely cover the field of translational platforms, we decided to also include commercial products in our review. We identified candidates through Google search and discussion with colleagues. The web search and discussions lead to the addition of one open-source platform and of three commercial products respecting the inclusion criteria. Overall, 11 platforms with advanced visualization capacities were included in the review: cBioPortal, interactive genomics patient stratification explorer (iGPSe), Igloo-Plot, G-DOC Plus, tranSMART, an unnamed data-cube-based model supporting heterogeneous data, Papilio, Caleydo Domino, Quicore Omics, Oracle Health Sciences Translational Research Center and OmicsOffice powered by TIBCO Spotfire. The first eight programs are open source, whereas the last three are commercial products.

We next identified the main features of each program analyzed along five major axes: general information, licensing, information content supported, visualization, and data exploration. This information was based on publicly available resources (i.e. original articles published in PubMed describing the systems and dedicated Web sites) and direct correspondence with the authors of the original papers or representatives for commercial products. In addition, we also include our personal experience using the program where available (based on using the five in-use open-source programs cBioPortal, Igloo-Plot, G-DOC Plus, tranSMART and Caleydo Domino as well as demo versions of Quicore Omics and OmicsOffice).

Results

Overview of multi-visualization tools

Our search results identified several flexible analytic tools or software programs with easy-to-use front-end graphic user interfaces (GUI) that have been developed to help researchers visualize complex data without needing deep data analytics or programming backgrounds. Tables 1 and 2 summarize general information, licensing, information content supported, visualization, and data exploration features for each system (Tables 1 and 2). The text below summarizes the systems in general with particular focus on visualization.

cBioPortal

cBioPortal, originally developed at the Memorial Sloan-Kettering Cancer Center, provides an interactive platform to visualize the data for over 120 different cancer studies [47, 48]. In a typical workflow, a researcher will accept a cancer study, select data type priority such as mutation and copy number alteration data, enter a list of genes of interest and then visualize various graphics summarizing the data slice. For example, researchers can investigate the frequency of specific mutations at each gene for the study, see scatter plots and box plots showing interaction between genomic events from different platforms and explore survival analyses where available. Advanced visualization features include an interactive Cytoscape graph that allows users to explore genes of interest within the larger network context and a MutationMapper graphic that allows interactive exploration population-wide genetic events linked to tables and three-dimensional (3D) visualizations. Some notable advantages of the tool are that it allows for easy integration with Integrative Genomics Viewer (IGV [49]) for more detailed genetic exploration and also provides a convenient REST-based web API (Application Programming Interface) that allows researchers an even wider range of analysis options. While the public online version is based on TCGA data sets, users can customize their instances by editing the code available through GitHub [50].

iGPSe

iGPSe is a proof of concept visual analytic system designed to allow users to perform complicated feature selection, clustering and subgroup comparison of genomic and clinical data without the need of deep programming or scripting knowledge [8]. Users begin by loading mRNA, microRNA (miRNA) and clinical data, as well as lists of genes of interest. The clustering analysis section allows patients to select clustering parameters and visualization results with heatmaps, silhouette plots and interactivity sparsity graphs. The final, integrative patient stratification, section contains interactive parallel sets based on clustering analysis linked to survival plots that allow real-time survival comparison of mRNA or miRNA clusters [51]. The principle advantage of this software was that, while applicable to other fields, it was developed with the input of domain experts in oncology to seamlessly integrate relevant features such as the various clustering algorithms, options to refine clusters and use of interactive summary pages.

Igloo-Plot

Igloo-Plot is an interactive visualization tool for multidimensional data in general developed by TATA Consultancy Services [52, 53]. Users download the application, upload their data according to predefined data formats and are presented with several, normalization, statistical analysis and clustering [54] and data visualization options. Options allowing for the selection of subgroups of samples or features are available through user-provided regular expressions. Principle visualization features include line graphs displaying variation across variables to aid in the normalization steps as well as the characteristic semicircular, or ‘igloo’ plot that facilitates the identification of clusters within the data and the identification of markers that define the clusters.

G-DOC Plus

G-DOC Plus is an updated version of the original G-DOC data management platform designed in 2011 to integrate structured clinical research with high-throughput data to advance precision medicine, translational research and population genetics [55, 56]. General visualization features include survival curves, Venn diagrams and heatmaps as well as those more specific for high-throughput analyses such as tools to visualize copy number instability, interaction networks and 3D representations of molecular targets. A principle feature of G-DOC Plus is its inherent comprehensive structure based on plug-ins to further its commitment to stay up-to-date with emerging omic technologies; the current version supports a wide variety of formats to accept mRNA, copy number variation, metabolite mass spectrometry and whole genome sequencing data. As of the date of manuscript drafting, G-DOC Plus allows users to explore data for >10 000 patients from over 50 public data sets from a wide variety of domains such as pediatric and adult oncology and wound healing. Data can also be loaded with the assistance of the support team by following a detailed data loading standard operating procedure.
| Category          | Item                                      | Freely available | Commercial            |
|-------------------|-------------------------------------------|------------------|-----------------------|
| Name of the platform | cBioPortal                               |                   | QIAGEN Omics Explorer |
|                   | iGPSe                                     |                   | Oracle Health Sciences |
|                   | Igloo-Plot                                |                   | Translational Research |
|                   | transSMART                                |                   | Center                 |
|                   | G-DOC                                    |                   | OmicsOffice powered   |
|                   | Papilio                                   |                   | by TIBCO Spotfire       |
|                    |                                           |                   |                       |
| Data housing       | MySQL                                     |                   |                       |
|                   | apache server                             |                   |                       |
|                   | Internal memory from loaded data          |                   |                       |
|                   | any Relational Database Management System |                   |                       |
|                   | (e.g. Oracle, PostgreSQL)                 |                   |                       |
|                   | Internal c++ data structures from data    |                   |                       |
|                   | Oracle                                   |                   |                       |
|                   | Internal memory from loaded data          |                   |                       |
|                   | SQLLite                                   |                   |                       |
|                   | SAS Cloud or on premis (MySQl)            |                   | Cloud or on premis     |
|                   | (Oracle or SQL)                           |                   | (Oracle or SQL)        |
| Principle frontend and/or backend programming languages | Java and Spring in back-end, Javascript, d3.js, R, perkTK |                   | Oracle ADF/Java        |
|                   | Javascript, d3.js, R, perkTK              |                   | EE on the front end,   |
|                   | Grails, Java, JavaScript                  |                   | with hooks into        |
|                   | C++ using a framework based on openGL and |                   | Oracle BI. The back-   |
|                   | C++                                       |                   | end in Oracle stack    |
|                   | Java, OpenGl/IOGL                         |                   | and middle tier        |
|                   | C++                                       |                   | so Oracle DB, Oracle   |
|                   |                                            |                   | BI/RF, Oracle          |
|                   |                                            |                   | Weblogic in a Java     |
|                   |                                            |                   | 2EE environment        |
|                   |                                            |                   |                       |
| Current status    | Dedicated domain                          |                   |                       |
|                   | In use                                    |                   |                       |
|                   | Exploration of large scale cancer         |                   |                       |
|                   | genomics sets                             |                   |                       |
|                   | In use                                    |                   |                       |
|                   | General visualization of multidimensional  |                   |                       |
|                   | datasets                                  |                   |                       |
|                   | In use                                    |                   |                       |
|                   | Hypothesis generation, hypothesis         |                   |                       |
|                   | validation, and cohort discovery in        |                   |                       |
|                   | translational research                    |                   |                       |
|                   | In use                                    |                   |                       |
|                   | Integrative analysis of various data types|                   |                       |
|                   | to uncover mechanisms                     |                   |                       |
|                   | In use                                    |                   |                       |
|                   | POC Exploration of heterogeneous data in  |                   |                       |
|                   | clinical cohorts                          |                   |                       |
|                   | In use                                    |                   |                       |
|                   | General visualization of multidimensional  |                   |                       |
|                   | datasets                                  |                   |                       |
|                   | In use                                    |                   |                       |
|                   | Visualization, exploration, and analysis   |                   |                       |
|                   | of bioinformatics data                     |                   |                       |
|                   |                                            |                   |                       |
| Licensing        | Software availability                     |                   |                       |
|                   | OpenSource GNU Affero General Public       |                   |                       |
|                   | License, version 3                         |                   |                       |
|                   | Open source                               |                   |                       |
|                   | Open source                               |                   |                       |
|                   | Open source                               |                   |                       |
|                   | NA                                        |                   |                       |
|                   | NA                                        |                   |                       |
|                   | Open-source (BSD License)                 |                   |                       |
|                   | Fee-based                                  |                   |                       |
|                   | Fee-based                                  |                   |                       |
|                   | Fee-based                                  |                   |                       |
| Client side       | interface                                 |                   |                       |
|                   | Web browser                               |                   |                       |
|                   | Web browser                               |                   |                       |
|                   | Stand-alone for Linux or Windows           |                   |                       |
|                   | Web browser                               |                   |                       |
|                   | Stand-alone stand-alone (Troltech Optimus) |                   |                       |
|                   | Web browser or standalone                 |                   |                       |
|                   | Stand-alone                               |                   |                       |
|                   | Stand-alone                               |                   |                       |
| User mailing list or support | No                                        |                   | Yes                    |
|                   | No                                        |                   | Yes                    |
|                   | No                                        |                   | Yes                    |
|                   | No                                        |                   | Yes                    |
|                   | No                                        |                   | Yes                    |
|                   | No                                        |                   | Yes                    |

Table 1. Summary of visualization programs for multidimensional data that can be applied to user-provided datasets. For each tool reviewed, we evaluated a number of features organized by the various categories: General Information, Licensing. PoC — Proof of Concept.
| Category          | Subcategory       | Item                                      | Freely available | Commercial |
|------------------|-------------------|-------------------------------------------|------------------|------------|
| General          | General information | Name of the platform                      | FREELY AVAILABLE | COMMERCIAL |
| Information content supported | Clinical Demographics | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Biology Survival  | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Imaging           | No | No | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes |
|                  | Genes              | Yes | Yes | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes |
| Omics            | mRNA              | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
|                  | Methylation, protein and phosphoprotein data | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Other            | Any type of raw or processed data that corresponds in a one-to-one relation to a sample | No | No | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes |
| Visualization    | High dimensional Correlation matrix | Yes (through IGV) | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes |
|                  | Parallel coordinates | No | No | No | No | No | No | No | No | No | No | No | No |
|                  | Scatterplot       | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Kaplan-Meier survival plot | Yes | No | No | No | No | No | No | No | Yes | Yes | Yes | Yes |
|                  | Kaplan-Meier survival plot | Yes | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Bar chart          | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Pie chart          | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Other              | MutationMapper, volcano plot | No | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| Data-exploration | Statistics and data mining | Yes | Log-rank test, P value, k-means, spectral clustering and community detection | No | No | No | No | No | No | No | No | No | No | No |
|                  | Limited views      | Yes | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Statistics         | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Logistic regression, correlation, t-test, ANOVA, basic summary statistics, hierarchical clustering, k-means clustering | No | No | No | No | No | No | No | No | No | No | No | No |
|                  | PCA, differential expression analysis, hierarchical clustering, group comparisons | No | No | No | No | No | No | No | No | No | No | No | No |
|                  | Yes | Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided by R in future versions | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|                  | Basic statistics such as finding differences in measures between two groups, confidence-weighted principal component ellipses | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

ANOVA = analysis of variance.
TranSMART

TranSMART is a rapidly growing web-based robust research management and analysis platform based on N-tier (data, business, presentation tiers in this case) architecture and Java schema designed to integrate disparate data sources to close the gap between basic science and clinical practice currently used by >100 organizations around the world. It features a simple user interface involving drag-and-drop movements that allows for an interactive analysis of a wide variety of data (demographic, diagnosis, medication, genetic, etc.) [57, 58] (Figure 2). The default installation provides a wide variety of basic, noninteractive, R-based plotting options such as scatterplots, bar charts, histograms, as well as

Figure 2: Overview of tranSMART. In a typical workflow, users define subsets of patients based on a drag and drop method of variables from the right column to the appropriate boxes (A). In this example, the summary statistics view (B) shows age difference between patients with genotypes (subsets 1 and 2, respectively) in a candidate gene. A colour version of this figure is available at BIB online: https://academic.oup.com/bib.
more complex waterfall plots, Manhattan plots and frequency plots for genomic analysis. TranSMART benefits from a growing worldwide community dedicated to improving its data processing and analytic features as well as its visualization features. For example, one project in our group involves the expansion of visualization capabilities of a plug-in called SmartR, a grails plug-in designed to improve the visual analytics tranSMART through advanced visualization libraries such as d3.js [59].

**Data-cube-based model supporting heterogeneous data**

The next tool in which we were interested was a proof of concept developed by Angelelli et al. [6] based on a data-cube-based model and designed for the visual exploration and analysis of large heterogeneous medical cohort studies. This software allows researchers to upload various data sets such as radiology results and cognitive scoring, slice patient groups based on specific features and then visualize how the data correlate with each other. The principle visualization component consists of a multiple-view dashboard featuring scatterplots, histograms and a 3D brain atlas color-coded by fiber bundle. These visualizations are all coordinated with each other based on interactive drag and drop or highlighting functions that allow users to select variables or data points of interest. The main advantages of this system are the flexibility of accepting incomplete, partial overlapping data reflective of real-world situations as well as the structure of the data storage, which allow fast, flexible calculations describing the relationships between different pieces of data.

**Papilio**

Papilio is another interactive tool that leverages visual analytics developed to explore heterogeneous medical cohort data to guide medical researchers and facilitate hypothesis generation, especially when no evident hypotheses are initially favored. After loading data, a first module called PrePap prepares the data. Next, the visualization module, VisPap, offers an interactive data exploration environment where users interact with a dashboard showing scatterplots, parallel coordinates and line diagrams all coordinated so as to maintain relationships and dependencies of data. Users also have the ability to visualize statistical analyses such as confidence-weighted principal component ellipses overlaid onto the data. Its principle features include a thorough image-processing pipeline that prepares raw images for downstream analysis as well as its robust conceptual framework based on domains, features and mappers that enhance the flexibility of the database while maintaining relationships between data.

**Caleydo domino**

Domino is a flexible data-visualization tool that improves the extraction, manipulation and comparison of interconnected heterogeneous subsets of multidimensional data sets in general [60, 61]. Users position draggable blocks in a workspace to rapidly assemble complex coordinated graphical schema representing the data and relationships between subsets. The software features a wide variety of simple and complex visualizations to incorporate into the schema ranging from histograms and scatterplots to parallel coordinate plots, mosaic plots and Sankey diagrams [62] (Figure 3). Two principle features include an intuitive GUI featuring placeholders and live previews that indicate possible drop locations and possible visualization to use as well as its library of innovative visualization techniques such as flexible linked axis (Figure 3).
multivariate data visualization”) and StratomeX, used for interactive visualization in cancer subtype analysis [64] (Figure 4).

**Qlucore omics**

As we believe, it is important to survey the widest variety of visualizations used to promote translational research using multidimensional data sets, we decided to additionally review available commercial solutions, the first of which is Qlucore Omics, a platform started in 2007 in Lund, Sweden optimized to explore biological data sets through interactive analysis and visualization features [66]. Data are loaded using a wizard, pre-processed and analyzed using a GUI workspace where users can select data and specific graphics and analyses to perform. The wide assortment of visualization supported range from scatter-plots and histograms to heatmaps and network visualizations all based on data and parameters selected from a tool bar. Users additionally have options to annotate data by features or statistics results, specify specific data or data slices to be plotted and synchronize visualizations such as by color codes to meet specific requirements. Like most commercial products, the software comes with complete documentation, support and comprehensive tutorials. An advantage of this program is the sheer amount of features available including calculations ranging from simple t-test statistics to advanced machine learning classifier builders.

**Oracle health sciences translational research center**

Oracle Health Sciences translational research center (TRC) provides a standardized industrial architecture that helps store, integrate and analyze multi-omic and clinical data and is specifically designed to facilitate biomarker discoveries, validation and application to clinical care [67]. The software’s top layer component is a cohort explorer used to identify and stratify clinical cohorts based on various normalization and filtering criteria. A principle advantage of the system is that it contains a rich omics data bank compiled from a large number of public studies that helps fit the project at hand into the context of up-to-date literature as well as promote cross-study omics data analysis. Of note, while the TRC supports direct integration with statistical and visualization software or even natural language processing functionality for test reports, these features are not included in the basic system package.

**Omicsoffice® powered by TIBCO Spotfire**

Our final commercial product to review is OmicsOffice, a comprehensive genomics data analysis tool backed by the TIBCO Spotfire data visualization and analytics software [68, 69]. Users work almost entirely within the GUI environment to perform genomic experiments and analyze data with almost no data preprocessing required start to finish. Visualization is based on a coordinate dashboard view where users can visualize all graphs and data as well as choose which data are displayed in real time using mouse-guided data slicing features. Visualization techniques span the gamut ranging from interactive bar and pie charts to pathway viewers and volcano plots for genomic results. OmicsOffice recognizes a wide range of proprietary omics data formats and includes workflows for integrating and running group comparisons on cross-platform data. Several benefits of the program are the comprehensive, peer-reviewed ‘click and go’ analytic pipelines for specific experiments such as quantitative polymerase chain reaction (qPCR),

---

Figure 4: A demonstration of StratomeX using exploration of a set of multiple tabular data sets for the TCGA clear cell renal carcinoma data set. This figure displays the main user interface of the program where users can drag and position data subsets and chose which calculations or visualizations to use to explore data and relationships between data. Above, users can visualize the relation between patients with subtypes based on two different genomic clustering experiments [65]. A colour version of this figure is available at BIB online: https://academic.oup.com/bib.
microarrays and NGS that take in raw data and produce full reports containing publication-ready graphics and information on quality control.

Discussion

In this manuscript, we have provided a detailed review investigating current visualization tools for multidimensional, big clinical research data sets used to promote translational research. We believe thorough visualization that integrates diverse data sources will become increasingly relevant in an environment where digitalization of the health field continues to accelerate.

Limitations

For the purpose of this review, we limited the scope to platforms controlled by intuitive graphical user interfaces that were flexible in receiving user-provided data. However, one related area that could have implications for visualization in translational research in general are tools developed to investigate data from fixed input data sets, usually arising from large multi-institutional research studies consisting of various data from hundreds or thousands of patients. In addition, we discuss additional techniques that have been used to visualize data in the medical field not limited to those used in the translational research applications we have described above.

Heterogeneity of the reviewed platforms

The use cases covered by the different platforms are heterogeneous (general cohort exploration, genomics analysis, general translational research and so forth). However, most of the systems could be used for a variety of applications leveraging similar data. Although the analytical capacities of platforms are complex to compare because of their difference in scope, we believe that the visualization features are relevant to explore together. In addition, we believe it was necessary to include visualizations from a variety of use cases to include the most comprehensive picture of contemporary visualization trends for exploration of heterogeneous health-related data sets.

Tools designed to visualize data for specific data sets

Data visualization has been shown to be especially helpful in oncology research where visualization is crucial for understanding certain genomic events, verifying data quality and identifying important aspects in cancer development (see [21] for thorough review). For example, NetGestalt [70] allows for multimodal exploration of the colorectal cancer TCGA data set and canEvolve [71] allows for integrated exploration of multiple TCGA studies. Note that while the current version of cBioPortal is dedicated primarily for the TCGA cancer data sets, we decided to keep this platform in our review because of its code availability and its strong presence in the translational research community. In addition, SysBioCube is an integrative data analysis platform designed by the US Army Medical Research group to study posttraumatic stress disorder [72], and Data Portal is a tool for interactive exploration of cognitive and radiological data for pediatric patients [73]. These tools allow researchers to intuitively explore rich data sets to uncover important biological pathways, regulation networks or drug targets.

Additional visualization techniques used in health research

A thorough review of emerging innovative visualization techniques for high-dimensional, complex data through innumerable ways of mapping of data variables to visual features such as position, size, shape and color is presented by Heer et al. [1]. For example, in visualizing time series data, various methods such as stacked graphs or index graphs showing percentage of change based on a selected point are available. Various techniques have been proposed to convert time data and events into optimal formats to facilitate quick interactive visualization [74, 75]. KNAVE-II is a tool designed to analyze and visualize time-oriented clinical data, whose principle feature is being able to classify and characterized raw time data using a predefined knowledge base [76]. In addition, a growing number of methods exist to represent spatial data such as color encoding (choropleth maps), overlaying graduated symbols or size distortion (cartograms). Spatial representation and cartography are also used in various medical research domains including brain function mapping [77], exploration of topographical distribution of skin molecules [78], identification of splice events in neurons [79] and of course the more traditional domain of epidemiology [80]. Finally, a number of graph methods have been used to visualize the relation between the different points in a network such as force-directed layouts, arc diagrams and, as discussed previously, matrix views. In medical research, network visualization is especially useful in exploration of genetic or proteomic information and molecular pathways [81, 82], and several tools exist to facilitate this process [83, 84].

Desiderata

Throughout our search of contemporary tools for multidimensional data visualization as approached from scientific domains, but also through additional searches spanning other domains where big data also poses challenges and opportunities such as data journalism, security and human-machine interface, we noticed several themes continually reemerging. Going forward, we believe that tools for multidimensional data visualization could be enhanced by adding capabilities for patient slicing, coordinated views, interactivity, flexibility, scalability and statistical power. We briefly describe each feature below.

Patient slicing, grouping or clustering

Multidimensional data sets with large numbers of samples or features are typically difficult to fully grasp by humans without some type of synthesis. As a result, various types of dimension reduction techniques such as principal component analysis (PCA) [85], self-organizing maps [86] and local linear embedding [87] have been proposed to simplify the data to only the most salient features. In addition, at the individual patient level, especially in studies with hundreds or thousands of patients, it is important to be able to select only relevant samples according to features or clusters of similar samples. This was important for our project consisting of data from a wide variety of sources and helped us, for example, separate out the effects of methylation (epigenetic) and genetic mutations for risk of transition to psychosis.

Coordinated or linked views

Moreover, visualization tools for multidimensional visualization are enhanced with multiple coordinated views, allowing users to see the same data set from different perspectives at once. This enables flexible exploration of various nuanced hypotheses with interactive data selection, or ‘brushing’, and can be applicable in a variety of domains outside of medicine from international politics to baseball [88]. Two interesting examples are PRISMA, which allows users to see uploaded data
In our study, it was important not only to group or cluster patients, but also to understand or measure the strength of the clusters or differences between them. It is, thus, important that any program we have will be backed by a powerful statistics package. We would like to acknowledge the clinical research team headed by Dr Marie-Odile Krebs for providing us with a complex clinical and biological data set that inspired our research for visualization methods.

Funding
BR has been funded in part by the CARPEM (Cancer Research for Personalized Medicine) research program.

References
1. Heer J, Bostock M, Ogievetsky V. A tour through the visualization zoo. Commun ACM 2010;53:59–67.
2. Fareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet 2011;52:413–35.
3. Leonelli S. Data interpretation in the digital age. Perspect Sci 2014;22:397–417.
4. Hey AJG, Tansley S, Tolle KM, et al. The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research, 2009. http://202.120.81.220:81/inter/uploads/reading/four-paradigm.pdf
5. Steenwijk MD, Milles J, Buchem MA. Integrated Visual Analysis for Heterogeneous Datasets in Cohort Studies. Eurographics Workshop Vis Comput Biomed, 2010. https://www.researchgate.net/profile/Johan_Reiber/publication/265438762_Integrated_Visual_Analysis_for_Heterogeneous_Datasets_in_Cohort_Studies/links/552c2f3f0cf21acb0920c495.pdf
6. Angelelli P, Paolo A, Steffen O, et al. Interactive visual analysis of heterogeneous cohort-study data. IEEE Comput Graph Appl 2014;34:70–82.
7. Peraksis ED, Van Dam J, Szalma S. How informatics can potentiate precompetitive open-source collaboration to jump-start drug discovery and development. Clin Pharmacol Ther 2010;87:614–16.
8. Ding H, Wang C, Huang K, et al. iGPSe: a visual analytic system for integrative genomic based cancer patient stratification. BMC Bioinformatics 2014;15:203.
9. Raghupathi W, Raghupathi V. Big data analytics in health-care: promise and potential. Health Inf Sci Syst 2014;2:3.
10. Bates DW, Saria S, Ohno-Machado L, et al. Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients. Health Aff. 2014. https://content.healthaffairs.org/content/33/7/1123.full
11. Simpao AF, Ahumada LM, Rehman MA. Big data and visual analytics in anaesthesia and health care. Br J Anaesth 2015;115:350–6.
12. Suresh S. Big data and predictive analytics: applications in the care of children. Pediatr Clin North Am 2016;63:357–66.
13. Janke AT, Overbeek DL, Kocher KE, et al. Exploring the potential of predictive analytics and big data in emergency care. Ann Emerg Med 2016;67:227–36.

14. Weinstein JN, Collisson EA, Mills GB, et al. Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45:1113–20.

15. Hudson TJ, Anderson W, Artez A, et al. International Cancer Genome Consortium. International network of cancer genome projects. Nature 2010;464:993–8.

16. Weiner MW, Veitch DP, Aisen PS, et al. 2014 Update of the Alzheimer’s Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 2015;11:e1–120.

17. Hall D, Huerta MF, McAuliffe MJ, et al. Sharing heterogeneous data: the national database for autism research. Neuroinformatics 2012;10:331–9.

18. O’Donovan MC. What have we learned from the Psychiatric Genomics Consortium. World Psychiatry 2015;14:291–3.

19. Thompson R, Johnston L, Taruscio D, et al. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med 2014;29(Suppl 3):S780–7.

20. Holten D. Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans Vis Comput Graph 2006;12:741–8.

21. Heinrich J, Weiskopf D. Continuous parallel coordinates. IEEE Trans Vis Comput Graph 2009;15:1531–8.

22. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. Genome Med 2013;5:9.

23. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res 2015;43:W566–70.

24. Deng W, Wang Y, Liu Z, et al. Heml: a toolkit for illustrating heatmaps. PloS One 2014;9:e111988.

25. Harter JM, Wu X, Alabi O'S, et al. Increasing the perceptual salience of relationships in parallel coordinate plots. Proc SPIE Int Soc Opt Eng 2012;8294:82940T.

26. Viau C, McGuffin MJ, Chiricota Y, et al. The FlowVisMenu and parallel scatterplot matrix: hybrid multidimensional visualizations for network exploration. IEEE Trans Vis Comput Graph 2010;16:1100–8.

27. Johansson J, Ljung P, Jern M, et al. Revealing structure within clustered parallel coordinates displays. In: IEEE Symposium on Information Visualization, 2005. INFOVIS 2005. IEEE, pp. 125132.

28. Ying-Huey F, Ward MO, Rundensteiner EA. Hierarchical parallel coordinates for exploration of large datasets. Proceedings Visualization99:43–50. In: (Cat. No.99CB37067). IEEE.

29. Nguyen LK, Degasperi A, Cotter P, et al. DYVIPAC: an integrated analysis and visualisation framework to probe multidimensional biological networks. Sci Rep 2015;5:12569.

30. Castro-Vega IJ, Letouzé E, Burnichon N, et al. Multi-omics analysis defines core genomic alternations in pheochromocytomas and paragangliomas. Nat Commun 2015;6:6044.

31. Arakawa K, Tomita M. Merging multiple omics datasets in silico: statistical analyses and data interpretation. Methods Mol Biol 2013;985:459–70.

32. Meng C, Kuster B, Culhane AC, et al. A multivariate approach to the integration of multi-omics datasets. BMC Bioinformatics 2014;15:162.

33. Hu Y, Yan C, Hsu C-H, et al. OmicCircos: a simple-to-use R package for the circular visualization of multidimensional omics data. Cancer Inform 2014;13:13–20.

34. Martin TC, Yet I, Tsai P-C, et al. coMET: visualisation of regional epigenome-wide association scan results and DNA methylation patterns. BMC Bioinformatics 2015;16:131.

35. Zhang H, Meltzer PS, Davis SR. caOmicSv: an R package for visualizing multidimensional cancer genomic data. BMC Bioinformatics 2016;17:141.

36. Ripley BD. The R project in statistical computing. MSOR Connections The newsletter of the LTSN Maths. 2001. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.449.6899&rep=rep1&type=pdf

37. matplotlib: python plotting — Matplotlib 1.5.1 documentation. http://matplotlib.org (9 May 2016, date last accessed).

38. Interactional JavaScript charts for your webpage | Highcharts. http://www.highcharts.com (9 May 2016, date last accessed).

39. Chart.js | Open source HTML5 Charts for your website. http://www.chartjs.org/ (9 May 2016, date last accessed).

40. Dygraphs | dygraphs is a fast, flexible open source JavaScript charting library. http://dygraphs.com (9 May 2016, date last accessed).

41. Belmonte NG. Javascript infovis toolkit. Web (Ed), 2011. http://thejit.org.

42. Bostock M, Ogievetsky V, Heer J. D3: data-driven documents. IEEE Trans Vis Comput Graph 2011;17:2301–9.

43. JavaScript (HTML5) Charting Library Comparisons. http://www.fusioncharts.com/javascript-charting-comparison/ (9 May 2016, date last accessed).

44. Kouskoumvekaki I, Shublaq N, Brunak S. Facilitating the use of large-scale biological data and tools in the era of translational bioinformatics. Brief Bioinform 2014;15:942–52.

45. Canuel V, Rance B, Avillach P, et al. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. Brief Bioinform 2015;16:280–90.

46. US National Library of Medicine National Institutes of Health. PubMed. NCBI. http://www.ncbi.nlm.nih.gov/pubmed (9 May 2016, date last accessed).

47. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cbioPortal. Sci Signal 2013;6:11.

48. cbioPortal for Cancer Genomics. http://www.cbioportal.org/ (9 May 2016, date last accessed).

49. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.

50. cbioPortal. cbioPortal/chiportal. GitHub. https://github.com/cbioPortal/chiportal/ (9 May 2016, date last accessed).

51. tindinghao. iGPSe. Youtube. 2014. https://www.youtube.com/watch?v=j8tCxOLjHg (accessed 8 July 2016, date last accessed).

52. Kuntal BK, Ghosh TS, Mande SS. Igloo-Plot: a tool for visualization of multidimensional datasets. Genomics 2014;103:11–20.

53. Igloo-Plot. http://metagenomics.atc.tcs.com/iglooPlot/ (9 May 2016, date last accessed).

54. Nováková L, Stepánková O. Multidimensional Clusters in RadViz. In: Proceedings of the 6th WSEAS International Conference on Simulation, Modelling and Optimization. Stevens Point, WI: World Scientific and Engineering Academy and Society (WSEAS), pp. 47075.

55. Madhavan S, Gusev Y, Harris M, et al. G-DOC: a systems medicine platform for personalized oncology. Neoplasia 2011;13:771–83.

56. Bhuvaneshwar K, Belouali A, Singh V, et al. G-DOC Plus - an integrative bioinformatics platform for precision medicine. BMC Bioinformatics 2016;17:193.

57. Scheufele E, Arzonon D, Coopersmith R, et al. tranSMART: an open source knowledge management and high content data
et al
60. Gratzl S, Gehlenborg N, Lex A, et al. Domino: extracting, comparing, and manipulating subsets across multiple tabular datasets. IEEE Trans Vis Comput Graph 2014;20:2023–32.

61. Home | Caleydo. http://www.caleydo.org/ (9 May 2016, date last accessed).

62. Riehmann P, Hanfler M, Froehlich B. Interactive Sankey diagrams. Inf Vis 2005;3:233–40. In: IEEE Symposium on INFOVIS 2005.

63. Caleydo Project. Domino: extracting, comparing, and manipulating subsets across multiple tabular datasets. Youtube 2014. https://www.youtube.com/watch?v=bsm59YQYbaQ (9 May 2016, date last accessed).

64. Lex A, Streit M, Schulz H-J, et al. StratomeX: visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization. Comput Graph Forum 2012;31:1175–84.

65. Caleydo Project. Guided visual exploration of genomic stratifications in cancer. Youtube 2014. https://www.youtube.com/watch?v=s2Zodj2GvHU (9 May 2016, date last accessed).

66. Qlucore - The D.I.Y Bioinformatics Software. http://www.qlucore.com/ (9 May 2016, date last accessed).

67. Oracle Health Sciences Translational Research Center - Overview | Oracle. http://www.oracle.com/us/products/applications/health-sciences/translational-research/index.html (9 May 2016, date last accessed).

68. Data Visualization & Analytics Software - TIBCO Spotfirehttp://spotfire.tibco.com/ (9 May 2016, date last accessed).

69. OmicsOffice® | Integromics. https://www.integromics.com/omicsoffice-suite/(9 May 2016, date last accessed).

70. Zhu J, Shi Z, Wang J, et al. Empowering biologists with multi-omics data: colorectal cancer as a paradigm. Bioinformatics 2015;31:1436–43.

71. Samur MK, Yan Z, Wang X, et al. canEvolve: a web portal for integrative oncogenomics. PLoS One 2013;8:e56228.

72. Chowbina S, Hammamieh R, Kumar R, et al. SysBioCube: A Data Warehouse and Integrative Data Analysis Platform Facilitating Systems Biology Studies of Disorders of Military Relevance. AMIA Jt Summits Transl Sci Proc 2013;2013:34–8.

73. Bartsch H, Thompson WK, Jernigan TL, et al. A functional cartography of cognitive systems. PLoS Comput Biol 2015;11:e1004533.

74. Bouslimani A, Porto C, Rath CM, et al. Molecular cartography of the human skin surface in 3D. Proc Natl Acad Sci USA 2015;112:E2120–9.

75. Treutlein B, Gokce O, Quake SR, et al. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. Proc Natl Acad Sci USA 2014;111:E1291–9.

76. Galanaud P, Galanaud A, Giraudoux P. Historical Epidemics Cartography Generated by Spatial Analysis: Mapping the Heterogeneity of Three Medieval ‘Plagues’ in Dijon. PLoS One 2015;10:e0143866.

77. Bouslimani A, Porto C, Rath CM, et al. Molecular cartography of the human skin surface in 3D. Proc Natl Acad Sci USA 2015;112:E2120–9.

78. Theocharidis A, van Dongen S, Enright AJ, et al. Network visualization and analysis of gene expression data using BioLayoutExpress3D. Nat Protoc 2009;4:1535–50.

79. Turkarslan S, Wurttmann EJ, Wu W-J, et al. Network portal: a database for storage, analysis and visualization of biological networks. Nucleic Acids Res 2014;42:D184–90.

80. Salama CR, Keller M, Kohlmann P. High-level user interfaces for transfer function design with semantics. IEEE Trans Vis Comput Graph 2006;12:1021–8.

81. Ce´ol A, Verhoef LGGC, Wade M, et al. Aire-la-Ville, Switzerland: Eurographics Conference on Visualization. Eurographics Association, pp. 131–8.

82. Zhao X, Kaufman A. Multi-dimensional reduction and transfer function design using parallel coordinates. Vol Graph 2010;69–76.

83. Weaver C. Multidimensional visual analysis using cross-filtered views. In: 2008 IEEE Symposium on Visual Analytics Science and Technology. IEEE, 2008, pp. 163–70.

84. Godinho PIA, Meiguins BS, Meiguins ASG, et al. PRISMA - a multidimensional information visualization tool using multiple coordinated views. In: 2007 11th International Conference Information Visualization (IV ’07). IEEE, 2007, pp. 23–32.

85. Gribov A, Sill M, Lück S, et al. SEURAT: visual analytics for the integrated analysis of microarray data. BMC Med Genomics 2010;3:21.

86. Doleisch H, Gasser M, Hauser H. Interactive feature specification for focus+ context visualization of complex simulation data. In VisSym, pp. 239–48.

87. Kaisler S, Armour F, Espinosa JA, et al. Big data: issues and challenges moving forward. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), pp. 995–1004.

88. Schumacher A, Rujan T, Hoefkens J. A collaborative approach to develop a multi-omics data analytics platform for translational research. Appl Transl Genom 2014;3:105–8.

89. Ooms J. The OpenCPU System: Towards a Universal Interface for Scientific Computing through Separation of Concerns. arXiv stat.CO. 2014. http://arxiv.org/abs/1406.4806

90. Horner J. RApache: Web application development with R and Apache.