

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, Qucore Omics, Oracle Health Sciences Translational Research Center and OmicsOffice[®] 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

William Dunn, Jr., BS, is a second-year Masters of science student in biomedical informatics at the Paris Descartes University and has a strong interest in translational science technologies and data visualization.

Anita Burgun, MD, PhD, is the head of the 'Information Sciences to support Personalized Medicine' group at the Centre de Recherche des Cordeliers, and CMIO at the European Hospital Georges Pompidou AP-HP Paris.

Marie-Odile Krebs, MD, PhD, among other roles, is president of the Clinical Research Committee, team leader in the Center for Psychiatry and Neurosciences and Head of department in the Sainte-Anne Hospital. Her expertise lies in exploring mechanisms of major mental health problems.

Bastien Rance, PhD, is a postdoctoral researcher working on clinical and omics data integration for the Cancer Research and Personalized Medicine project. His expertise is in data integration, data mining and controlled terminologies.

Submitted: 16 May 2016; **Received (in revised form):** 30 July 2016. **Accepted:** 3 August 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

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, Matlab®, 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 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], Dygraphs [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

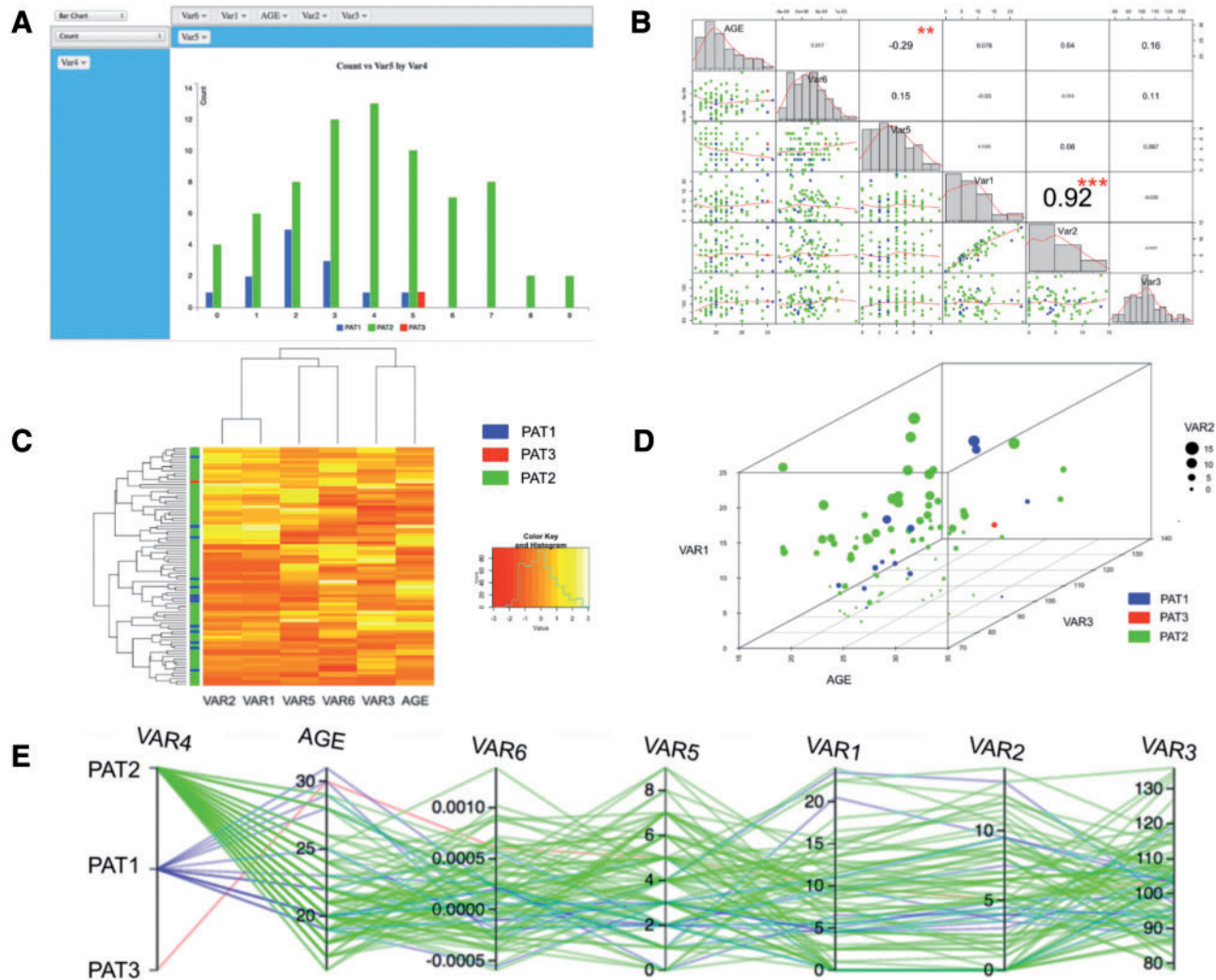


Figure 1. A sampling of commonly used visualization techniques for multidimensional data using a subset of data in our data set compiling data from three groups of patients. Var1, Var2 and Var3 are neurocognitive dimensions, Var4 and Var5 are psychopathological dimensions and Var6 is a global genetic index. Specific visualizations used are (A) dynamic pivot table (using R 'rpivotTable' 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>.

library are easy to leverage for a bioinformatician, but the knowledge gap is often too wide for biologists and clinicians without a background in bioinformatics or biostatistics. A common challenge 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 clinical and omics data to advance health science and translational research.

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 clinical and omics data), (ii) that feature data visualization functionalities 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 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 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 transSMART]. We then searched scientific 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, details are available in [Supplementary Table S1](#)). We identified three new platforms through this step, and one from citations

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, Qlucore 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 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 Qlucore 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.

Table 2. 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: Information content supported, visualization, data-exploration. PoC = Proof of Concept.

Category	Sub-category	Item	Freely available											
General information	General information	Name of the platform	cbioPortal	iGPSe	Igloo-Plot	transSMART	G-DOC Plus	Data-cube-based model supporting heterogeneous data	Papilio	Caleydo Domino	QiLucore Omics Explorer	Oracle Health Sciences Translational Research Center	OmicsOffice® powered by TIBCO Spotfire	
Information content supported	Clinical	Demographics	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
		Diagnosis	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
		Biology	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
		Survival	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Omics	Imaging	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes		
	Gene mutation	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes		
	mRNA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes		
Other	Methylation, protein and phosphoprotein data	Methylation, protein and phosphoprotein data	miRNA	NA	NA	NA	NA	NA	NA	NA	Methylation, protein expression, flow cytometry	Methylation	RNA sequence, chromatin immunoprecipitation sequence, qPCR	
	Any type of raw or processed data that corresponds in a one to one relation to a sample	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	
Visualization	High dimensional	Heatmap	Yes (through IGV)	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	
		Correlation matrix	No	No	Yes	No	No	Yes	No	Yes	No	no	Yes	
Data-exploration	Statistics and data mining	Parallel coordinates	No	No	No	No	No	No	Yes	Yes	No	no	Yes	
		Other	OncoPrinter	Parallel sets, silhouette plot, Sankey plot, force-directed graphs	NA	Waterfall plot, PCA plot, Haploview, Manhattan plot, Forest plot, Frequency plot for aCGH	Biological network and pathways viewers (Reactome, Cytoscape), integrated genome browser (Browse)	NA	Scatterplots color coded by patient type overlaid with PCA ellipses	Parallel sets, sankey-diagrams, and more novel graphics	Sample PCA, variable PCA	Requires business intelligence layer for visualization	Pathway viewer, 3D scatterplot, map chart, treemap	
		Timeline/line chart	No	No	No	Yes	No	No	Yes	No	Yes	Yes	Yes	
		Histograms	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	
		Scatterplots	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	
		Kaplan-Meier survival plot	Yes	Yes	No	Yes	Yes	No	No	No	Yes	Yes	No	
		Bar charts/box and whisker	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	
		Pie charts	No	Yes	No	Yes	no	no	no	No	Yes	no	Yes	Yes
		Other	MutationMapper, volcano plot	NA	NA	Novel semi-circle plotting approach based on correlation and Hooke's law	NA	Interactive 3D molecular viewer, chromosome and CNV visualizations, Venn diagram	Atlas view representing areas of brain implicated in analyses	NA	NA	NA	NA	Volcano plot
		Coordination	Statistics and data mining	Survival log-rank test, Cytoscape graph viewer for genetic networks	No	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
Log-rank test, k-means, spectral clustering and community detection	Yes			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	
Linked views	No			Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	
Statistics	Yes			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	
Survival log-rank test, Cytoscape graph viewer for genetic networks	No			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	
Log-rank test, k-means, spectral clustering and community detection	Yes			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	
Linked views	No			Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Statistics	Yes			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	
Survival log-rank test, Cytoscape graph viewer for genetic networks	No			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	
Log-rank test, k-means, spectral clustering and community detection	Yes			Yes	No	Logistic regression, correlation, t-test, Fisher test, ANOVA, basic summary statistics, hierarchical clustering	PCA, differential expression analysis, hierarchical clustering, group comparisons	Correlation statistics between radiology results and cognitive testing, multivariate statistics, multilinear regression, as well as any type of statistics provided calculated by R in future versions	Basic statistics such as finding differences in measures between two groups. Confidence-weighted principal component ellipses	NA	T-test, ANOVA, linear regression, quadratic regression, rank regression, classifier building and training SVM, RT, RNN	Integrated with programming languages such as R for statistics beyond simple group counts	Line similarity, regression modeling, wide range of parametric and nonparametric statistical tests, functional gene analysis, data classification	

ANOVA = analysis of variance.



Figure 2: Overview of TransSMART. 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>.

TransSMART

TransSMART 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

more complex waterfall plots, Manhattan plots and frequency plots for genomic analysis. TransSMART 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 transSMART 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 ('Flexible linked axes for

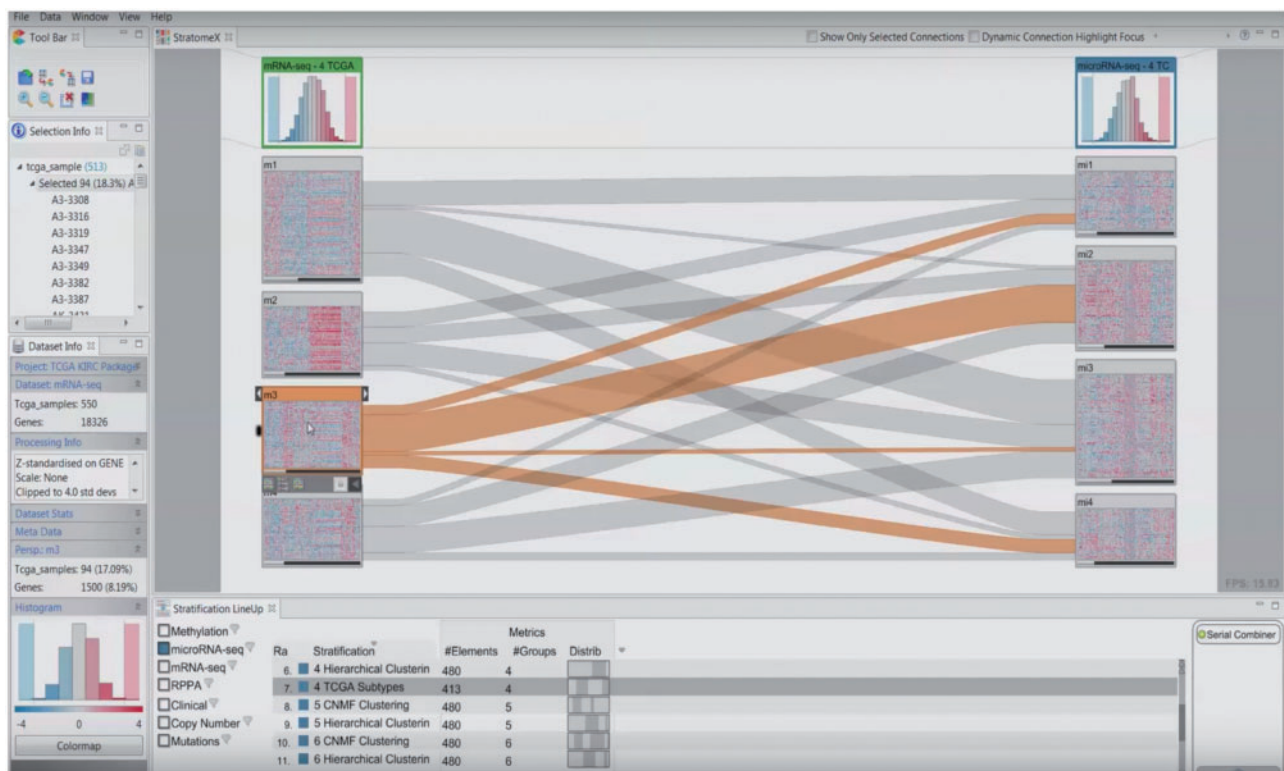


Figure 3: A demonstration of Caleydo Domino using exploration of a set of multiple tabular data sets for a music data set containing song and musician information. This figure displays the main user interface of the program where users can drag and position data subsets and choose which calculations or visualizations to use to explore data and relationships between data [63]. A colour version of this figure is available at BIB online: <https://academic.oup.com/bib>.

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),

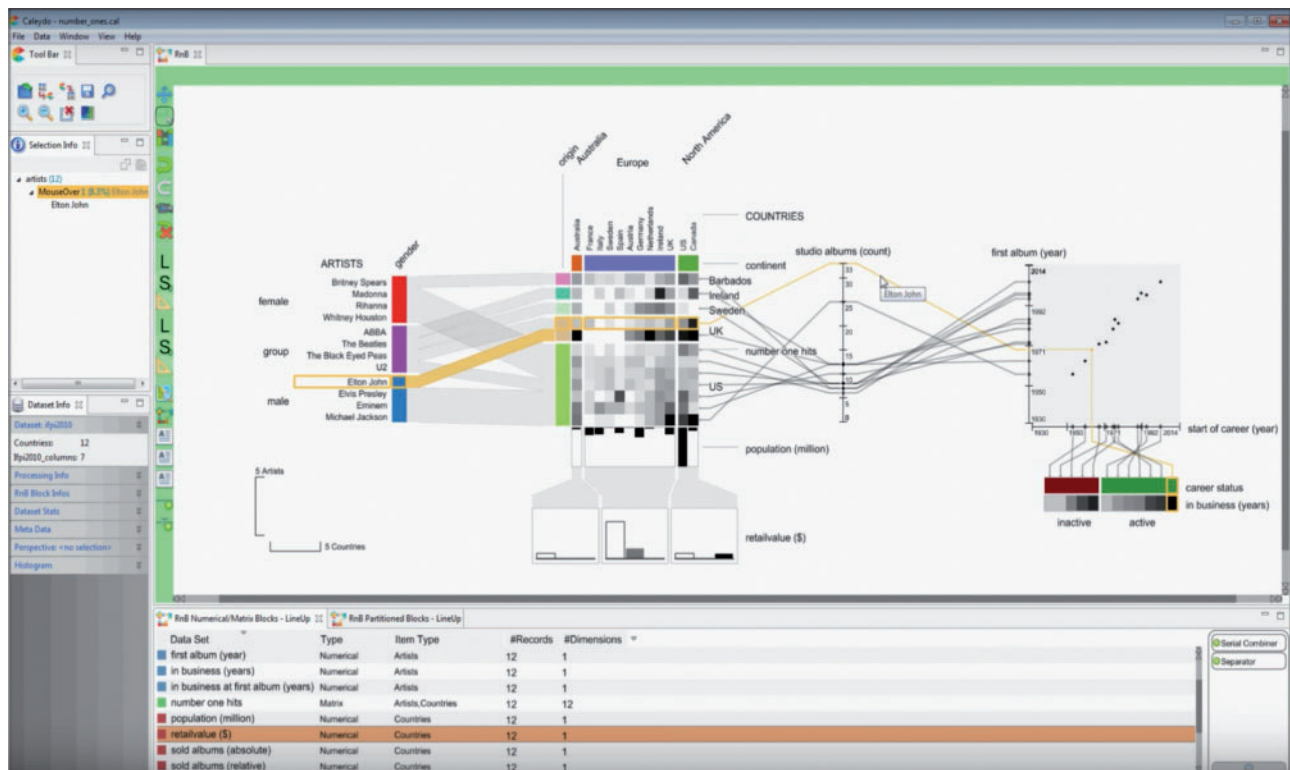


Figure 4: A demonstration of StratomeX using exploration of a set of multiple tabular data sets for the TCGA clear cell renal carcinoma data set. This figure displays the main user interface of the program where users can drag and position data subsets and chose which calculations or visualizations to 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 multi-omic 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 characterize 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 neurexins [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

represented by treemaps, scatterplots and parallel coordinates, all coordinated with each other in terms of color, filter and selection [89], and SEURAT, which combines linked views with exploratory analyses for microarray data visualization [90].

Interactivity

Often going hand in hand with patient slicing or coordinate views, interaction is a key aspect of visualization tools that facilitates flexible searching and localizing of interesting features in a data set through intuitive commands [91]. Many of the popular visualization platforms mentioned in the introduction consist of or support user interaction ranging from tooltips on mouse hover/touch to triggering the reordering of data or other complex actions.

Flexibility

Like many research groups, we are constantly changing what types and formats of data we collect based both on changes within the scientific community and the types of patients that enter our research center. This ‘variability’ issue is likely the most important challenge in analyzing big data [92]. It is, thus, important that tools be flexible to accept data types from a wide range of sources. We also understand that this may pose a limit, as measures to increase flexibility to accept different data by widening acceptable parameters or formats may force us to decrease the level of specificity and, thus, detail for a data source.

Scalability

Given the increasing data generated everyday by high-throughput experiments and technologies, another feature typically required for successful translational research is scalability [93]. In addition, it is important for visualizations to be able to efficiently transition through scales of magnitude while keeping an appropriate data granularity. For example, features should be implemented that support ‘drilling down’, to find specific information about outliers from high-level visualizations [5].

Statistical power

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 the differences between them. It is, thus, important that any program we have would be backed by a powerful statistics package. Much progress has been made in this domain in the past few years allowing statistics packages such as R be easily integrated into third-party software such as Web sites (‘embedded scientific computing’—see OpenCPU [94], rApache [95]).

Conclusion

In this work, we have presented a comprehensive review of the current tools in use for visualization of complex, multidimensional data sets. As medical research shifts increasingly toward a more data-driven approach, this need to comprehensively visualize multivariate data will continue to grow, especially in health-care research settings. We believe our work will serve a wide variety of investigators performing similar research.

Key Points

- Thorough multidimensional visualization offers several benefits with potential implications in understanding disease and ultimately improving patient care.

- Translation research platforms in the clinical domain provide an ideal setting for a wide range of multidimensional visualization applications.
- In this work, we summarize the existing landscape of these types of tools currently used as well as provide our input on points to consider in advancing their development.

Supplementary Data

Supplementary data are available online at <http://bib.oxfordjournals.org/>.

Acknowledgements

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. Pareek 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/readings/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/552c2f30cf21acb0920c495.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. Perakslis ED, Van Dam J, Szalma S. How informatics can potentiate precompetitive open-source collaboration to jumpstart 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 health care: 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. HemI: a toolkit for illustrating heatmaps. *PLoS One* 2014;**9**:e111988.
25. Harter JM, Wu X, Alabi OS, 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 FlowVizMenu 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 Visualization* 99:43–508. 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 LJ, Letouzé E, Burnichon N, et al. Multi-omics analysis defines core genomic alterations 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://citeseex.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. Interactive 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. NCB. <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, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6.
50. cBioPortal. cBioPortal/cbioportal. GitHub. <https://github.com/cBioPortal/cbioportal/> (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, Štěpánková 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, Aronzon D, Coopersmith R, et al. transSMART: an open source knowledge management and high content data

- analytics platform. *AMIA Jt Summits Transl Sci Proc* 2014;2014:96–101.
58. Potenzzone R. tranSMART Foundation. <http://transmartfoundation.org/> (9 May 2016, date last accessed).
 59. transmart. transmart/SmartR. GitHub <https://github.com/transmart/SmartR> (9 May 2016, date last accessed).
 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. Riehmman P, Hanfler M, Froehlich B. Interactive Sankey diagrams. *Inf Vis* 2005;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=bm59Y8QYbAQ> (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=s2Zofj2GVHU> (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 Spotfire <http://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 web-portal for interactive data exploration, visualization, and hypothesis testing. *Front Neuroinform* 2014;8:25.
 74. Riss M. FTSPLOT: fast time series visualization for large datasets. *PLoS One* 2014;9:e94694.
 75. Henriques TS, Mariani S, Burykin A, et al. Multiscale Poincaré plots for visualizing the structure of heartbeat time series. *BMC Med Inform Decis Mak* 2016;16:17.
 76. Shahar Y, Goren-Bar D, Boaz D, et al. Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. *Artif Intell Med* 2006;38:115–35.
 77. Mattar MG, Cole MW, Thompson-Schill SL, et al. A functional cartography of cognitive systems. *PLoS Comput Biol* 2015;11:e1004533.
 78. 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.
 79. 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.
 80. 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.
 81. Céol A, Verhoef LGGC, Wade M, et al. Genome and network visualization facilitates the analyses of the effects of drugs and mutations on protein-protein and drug-protein networks. *BMC Bioinformatics* 2016;17(Suppl 4):54.
 82. Theocharidis A, van Dongen S, Enright AJ, et al. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nat Protoc* 2009;4:1535–50.
 83. Turkarlsan S, Wurtmann 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.
 84. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* 2011;696:291–303.
 85. 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.
 86. de Moura PF, Freitas CMDS. Design of multi-dimensional transfer functions using dimensional reduction. In: *Proceedings of the 9th Joint Eurographics/IEEE VGTC Conference on Visualization*. Aire-la-Ville, Switzerland: Eurographics Association, pp. 131–8.
 87. Zhao X, Kaufman A. Multi-dimensional reduction and transfer function design using parallel coordinates. *Vol Graph* 2010;69–76.
 88. Weaver C. Multidimensional visual analysis using cross-filtered views. In: *2008 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2008, pp. 163–70.
 89. 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.
 90. Gribov A, Sill M, Lück S, et al. SEURAT: visual analytics for the integrated analysis of microarray data. *BMC Med Genomics* 2010;3:21.
 91. Doleisch H, Gasser M, Hauser H. Interactive feature specification for focus+ context visualization of complex simulation data. In *VisSym*, pp. 239–48.
 92. 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.
 93. 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.
 94. 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>
 95. Horner J. RApache: Web application development with R and Apache.