In-Memory Technology Enables Interactive Drug Response Analysis

Matthieu-P. Schapranow* Konrad Klinghammer[†] Cindy Fähnrich* Hasso Plattner*

*Hasso Plattner Institute
Enterprise Platform and Integration Concepts
August–Bebel–Str. 88
14482 Potsdam, Germany
{schapranow|cindy.faehnrich|plattner}@hpi.de

[†] Charité – Universitätsmedizin Berlin
Comprehensive Cancer Center
Charitéplatz 1
10117 Berlin, Germany
konrad.klinghammer@charite.de

***Abstract*—Latest medical diagnostics generate increasing amounts of big medical data. Specific software tools optimized for the use by healthcare experts and researchers as well as systematic processes for data processing and analysis in clinical and research environments are still missing.**

Our work focuses on the integration of high-throughput next-generation sequencing data and its systematic processing and its instantaneous analysis to use them in the course of precision medicine.

We share our research results on designing a generic research process for drug response analysis including specific software tools built on top of our distributed in-memory computing platform for processing of big medical data. Furthermore, we present our technical foundations as well as process aspects of integrating and combining heterogeneous data sources, such as genome, patient, and experimental data.

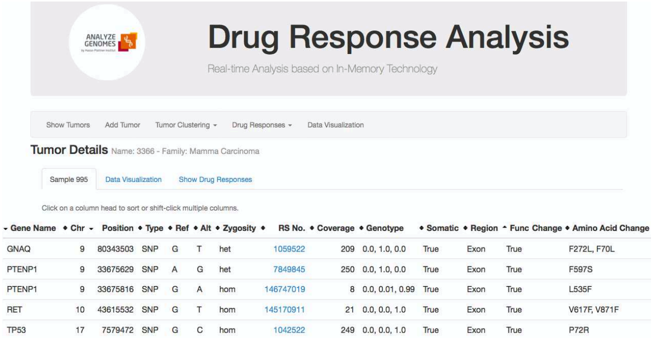
Keywords-Drug Response Analysis; Genome Data Analysis; Process Integration; In-Memory Database Technology; E-Health; Next-Generation Sequencing.

I. Introduction

The Human Genome (HG) project launched in the 1990s involved thousands of research institutes worldwide and required more than a decade to sequence and decode a single full HG [1]. Nowadays, Next-Generation Sequencing (NGS) technology are used to process genomes within hours at reduced costs and, thus, support innovative e-health applications, e.g., in course of precision medicine [2]. Precision medicine aims at treating patients specifically based on individual dispositions, e.g., genetic or environmental factors [3].

The In-Memory Database (IMDB) technology has proven to have major advances in analyzing big enterprise and medical data, e.g., to analyze patient data and identify pharmaceutical counterfeits in real time [4, 5].

In this work, we present our findings of applying IMDB technology to enable integration of experiment results, its real-time analysis, and prediction of drug response in silico in the course of precision medicine. We introduce a generic research process for cancer researchers built upon our High-performance In-memory Genome (HIG) cloud platform, which is online available at http://we.analyzegenomes.com/. The HIG platform provides cloud-based Internet services for processing and analysis of big medical data, such as high-throughput genome data. Together with cancer researchers, we developed special purpose software tools to evaluate results of conducted Xenograft experiments, combine them with relevant medical knowledge, and to analyze them instantaneously [5, 6]. Fig. 1 depicts a screenshot of our drug response cloud application showing details about genetic changes of a mama carcinoma tumor sample.



Tumor Details Name: 3366 - Family: Mamma Carcinoma

Click on a column head to sort or shift-click multiple columns.

Gene Name	Chr	Position	Type	Ref	Alt	Zygosity	RS No.	Coverage	Genotype	Somatic	Region	Func Change	Amino Acid Change
GNAQ	9	80343503	SNP	G	T	het	1059522	209	0.0, 1.0, 0.0	True	Exon	True	F272L, F70L
PTENP1	9	33675629	SNP	A	G	het	7849845	250	0.0, 1.0, 0.0	True	Exon	True	F597S
PTENP1	9	33675816	SNP	G	A	hom	146747019	8	0.0, 0.01, 0.99	True	Exon	True	L535F
RET	10	43615532	SNP	G	T	hom	145170911	21	0.0, 0.0, 1.0	True	Exon	True	V617F, V871F
TP53	17	7579472	SNP	G	C	hom	1042522	249	0.0, 0.0, 1.0	True	Exon	True	P72R

Figure 1. Screenshot of the drug response analysis cloud application built on our real-time analysis platform.

The rest of the paper is structured as follows: In Sect. II, our work is set in context of related work. We introduce selected in-memory technology building blocks in Sect. III and our research methodology in Sect. IV. We present the current drug response process in Sect. V and define our enhanced research process in Sect. VI. In Sect. VII, we evaluate our contribution while our work concludes with an outlook in Sect. VIII.

II. Related Work

Fig. 2 provides a comparison of costs for sequencing and main memory modules. Both costs follow a steadily declining trend, which facilitates the increasing use of NGS for whole genome sequencing and IMDB technology for its data analysis. Related work in the

978-1-4799-6644-8/14/$31.00 ©2014 IEEE

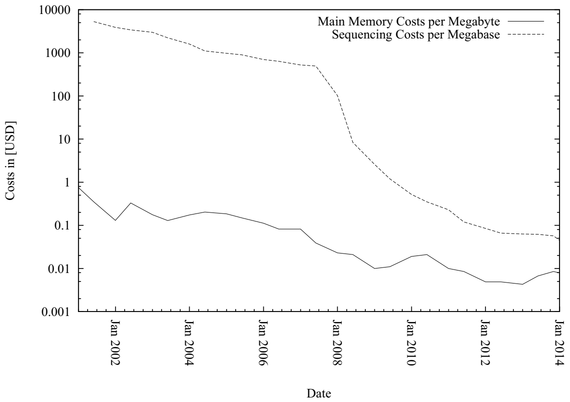


Figure 2. Costs for next-generation sequencing and main memory from 2001 to 2014 adapted from [7, 8].

field of e-health applications for genome data processing has increased in the recent years. However, work focusing on implementing end-to-end processes is still rare. Therefore, we focus on the implementation of innovative research processes by, amongst others, the integration of genome data processing and statistical data analysis in course of drug response analysis.

Sun investigated gene regulations in prostate cancer samples combining latest sequencing technology and bioinformatic approaches[9]. We agree that an integrated data processing and analysis approach is also essential for other application fields. Thus, we integrate various heterogeneous data sources to enable multi-modal modeling of diseases. Furthermore, we enable researchers for the first time to perform data analysis a) in real-time without any delay and b) without the need to involve dedicated IT experts, e.g., to prepare analysis reports.

Rossello et al. propose the use of Xenograft models as sources for preclinical work when primary tumor samples are rare, e.g., for small cell lung cancer. They share detailed insights into their methodology using state-of-the-art alignment and variant calling tools, such as BWA, GATK, and snpEff [11, 12, 13]. However, they do not provide a tight integration of their incorporated genome sequencing and data analysis pipeline, which consumed major parts of the their experimental time. Our contribution enables tight integration of experimental data, such as NGS tumor data, and its real-time analysis as described in Sect. V.

III. In-memory Technology Building Blocks

We refer to IMDB technology as a toolbox of IT artifacts enabling processing of data in real-time in the main memory of server systems [14]. The combination of IMDB database technology and analysis of genome data is driven by declining costs as described in Sect. II. In the following, we outline selected IMDB technology building blocks and their relevance for our work.

Insert-only: It is a data management approach that stores data changes as new entries. In contrast to traditional databases, in an insert-only database table operations that change data, such as update or delete, do no longer "destroy" the original data. Affected entries are invalidated instead while keeping the complete history of value changes accessible [14]. Thus, we guarantee a reproducible research process by tracing all data changes, e.g., to retrospectively analyze when experiment results were taken.

Lightweight Compression: It refers to a data storage representation, which consumes less space than its original pendant [14]. The columnar storage layout incorporated by our IMDB supports transparent lightweight compression techniques, such as run-length encoding, dictionary encoding, and difference encoding [15]. Typically, values of a database attribute are within a very small subset of the attribute's domain, e.g., `male` and `female` for the gender type. Lightweight compression maps all unique values to a uniform format, e.g., `male=1` and `female=2`, consuming the minimum portion of storage to represent the relevant subset of the complete data domain. Thus, lightweight compression reduces the amount of required main memory capacity to enable real-time analysis of big medical data.

Partitioning: We distinguish between vertical and horizontal partitioning [16]. The former refers to the arrangement of database columns. It is achieved by splitting columns of one database table in multiple column sets wile each set can be distributed on individual servers [17]. The latter addresses long database tables and their division into smaller chunks of data. Splitting data into equally long horizontal partitions supports parallel search operations and improves scalability [14]. In particular, we incorporate inter-chromosome, i.e., store data chromosome-wise, and intra-chromosome partitioning, i.e., store chromosome data in regions, to support data processing by individual CPU cores.

IV. Methodology

In the course of this project, we followed the design science methodology to improve the existing research process with the help of selected software artifacts [18]. For the development of the software tools, we applied the Design Thinking (DT) methodology. DT proposes to work in interdisciplinary teams with members from different disciplines, e.g., a software developer and a medical researcher [19]. They can combine their individual viewpoints on the problem domain while chances that important functional aspects are not recognized are minimized. Additionally, interdisciplinary teams will not suffer from rivalry between experts of the same field while having all required expertise to implement the

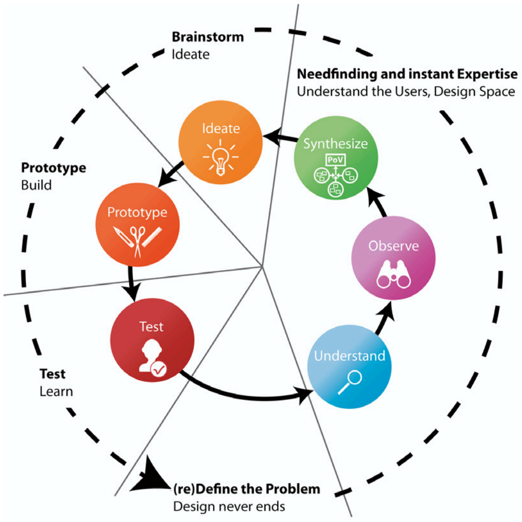


Figure 3. Adapted Design Thinking Process as defined by the HPI School of Design Thinking in Potsdam and Stanford [19].

solution available in the team. Furthermore, DT provides a process framework as depicted in Fig. 3 asking for constant communication between developing team, stakeholders, and targeted end users.

We incorporated DT as follows: we conducted user interviews with cancer researchers and physicians to document the existing research process as described in Sect. V. Furthermore, we designed an enhanced research process by combining data from heterogeneous sources and processing steps in a software prototype as described in Sect. VI. For rapid software development, we followed the scrum software development methodology [20]. Based on the constant user feedback, we extended our prototype iteratively following short development sprints of one to two weeks and evaluated new functionality either in workshops at users' site or telephone interviews where end users tested the software artifacts via screen sharing tools. The acquired feedback was incorporated to plan the next development sprint.

V. Current Drug Response Analysis Process

Nowadays, drug response analysis consists of a) conducting drug experiments, e.g., in Xenograft models, and b) the analysis of the obtained experiment results [21]. We observed that the following categories of data sources are used for drug response analysis as depicted in Fig. 4:

- **Patient Metadata** is retrieved from Clinical Information Systems (CISs) and contains specific patient details, such as age, gender, and anamnesis. Its data volume typically ranges from one to 100 MB excluding any diagnostic data, such as imaging data,
- **Genome Data** is obtained by sequencing resected tumor material, e.g., with NGS devices. Its data volume is in the range of some 100 MB for panel sequencing and up to 500 GB for NGS.
- **Experiment Data** is obtained by wet laboratory assistants, e.g., documenting the individual drug tests in Xenograft experiments. Its data volume is in the range from 10 MB to 1 GB.

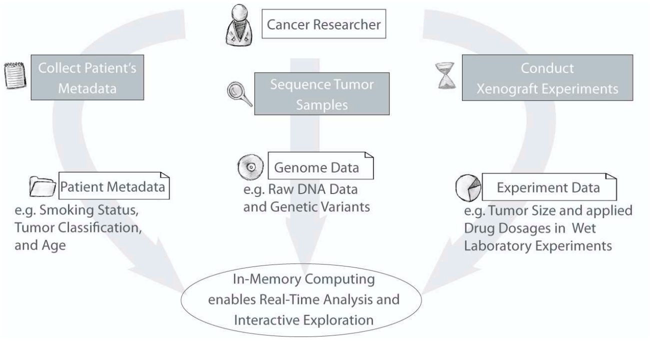

Cancer Researcher
Collect Patient's Metadata
Sequence Tumor Samples
Conduct Xenograft Experiments
Patient Metadata
e.g. Smoking Status, Tumor Classification, and Age
Genome Data
e.g. Raw DNA Data and Genetic Variants
Experiment Data
e.g. Tumor Size and applied Drug Dosages in Wet Laboratory Experiments
In-Memory Computing enables Real-Time Analysis and Interactive Exploration

Figure 4. The enhanced drug response analysis process involves data in heterogeneous formats from different data sources.

The time consumed for preparation and laboratory work can range from days to weeks depending on the conducted experiments. Although the data analysis phase is already assisted by software, it still takes days up to weeks to perform complex data analysis, such as correlation or cohort analysis. As a result, processing and analysis of data is the most time-consuming aspect of the research process after laboratory work. The reasons are many-fold, e.g., the absence of optimized tools for data analyses, analysis tools only tested for a small subset of required data sources, and time-consuming transformation of relevant data.

Manual or semi-manual time-consuming process steps, such as using spreadsheets for conducting complex data analysis, characterize all phases of the existing process. From a software engineering perspective, we focus on all process steps that involve digital data processing and analysis. Thus, our work focuses on the data processing and analysis of the existing research process to optimize the overall process performance and acceptance.

VI. Enabling Real-time Data Analysis using In-Memory Technology

Fig. 4 depicts our enhanced research process and the involved data sources. We applied in-memory technology for all data processing and analysis to improve the overall process performance. Our enhanced research process is divided in the following steps:

- **Computational biology** performs data processing, e.g., alignment of raw DNA,
- **Visual data exploration** supports verification of hypotheses by researchers, and
- **Clustering of tumor data** enables real-time classification of results.

A. Computational Biology

In the following, we share selected insights on enabling real-time computational processing of DNA data by incorporating IMDB technology.

1) Open Reading Frame Detection: Detecting new Open Reading Frames (ORFs) helps to find potential gene locations on the DNA [22].

Our ORF detection is two-divided: Firstly, we detect start and end codons in all reading frames. Secondly, pairs of start and end codons of the reading frame are analyzed to obtain ORFs of a minimum length.

We process the forward and the backward strand in parallel, e.g., when searching for the start codon "ATG" on the DNA's forward strand, the reserve-inverted triplet "CAT" to detect the start codon on the backward strand is also checked. The result consists of reading frame, position of the codon, and its type. These results are grouped by reading frame to identify the corresponding pairs of start and stop codons.

We implemented the ORF detection algorithm within the IMDB using SQLScript and L [23, 24]. Thus, we incorporate advances of in-memory computing, such as processing genome data directly within the IDMB eliminating former data transfer or transformations.

2) Detection of Genetic Functional Changes: For each genetic variant, its potential impact on the Amino Acid (AA), which is built from the genetic code, needs to be analyzed. Changes in the AA affect the proteins built from it, which might result in harmful mutations [25].

We implemented the detection of functional genetic changes as a stored procedure in the IMDB as follows: We join the variant's locus, i.e., chromosome and position, with a database table of known genes to determine whether the variant is located on a known gene [26]. If the variant is not located on a known gene, we consider its impact as minor since the current medical knowledge about variants not located on genes is very limited. If the variant is located on a gene, all splicing variants of the gene are processed in parallel to evaluate its impact per splicing variant.

We document an AA change using the expected followed by the detected AA including the position of the affected triplet, e.g., V600E describes an AA change of valine to glutamic acid at triplet 600.

B. Visual Data Exploration

We developed interactive visualization tools to enable researchers to perform interactive graphical exploration of data, wich are outlined in the following.

Hierarchical clustering creates a hierarchy of clusters by iteratively merging closest data points to a cluster (agglomerative hierarchical clustering). Thus, the clustering algorithm needs a measure of dissimilarity between sets of observations.

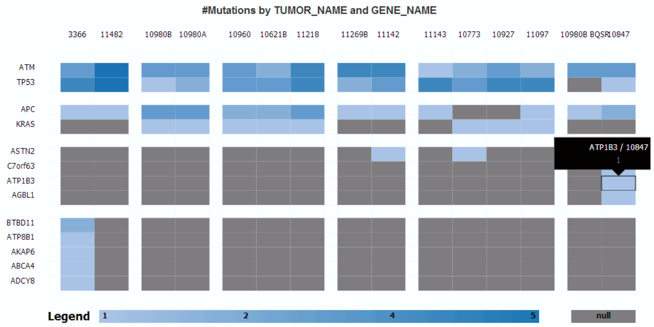


Figure 5. Clustered heat map using hierarchical clustering comparing mutation count, a subset of genes, and tumor samples.

Fig. 5 shows our clustered heat map visualizing the results of a hierarchical clustering.

For hierarchical clustering, the measure is formed by combining an appropriate metric for distance calculation between data points and a linkage criterion for calculating the distance between merged data points [32]. Rows and columns of the heat map are individually cluster using using row- and column-wise vectors. We used the Euclidean distance function to calculate the distance between vectors and singly linked as the linkage criteria.

Hierarchical clustering results in dendrograms, which represents nested clusters at certain levels of similarity. We use the dendrogram to rearrange/reorder the heat map, e.g., to mark gaps within the heat map.

C. Clustering of Tumor Data

In the following, we share our process enhancements to enable interactive classification of research data.

1) Tumor Data Association Rules: Association Rules Mining (ARM) requires a set S of item sets S_i as its data basis: $S = \{S_1, ..., S_m\}$. Every item set S_i consists of several items i_i from the list I of distinct items: $I = \{i_1, ..., i_n\}$. Item sets are processed to detect rules of type: $A \Rightarrow B$ where $A \subseteq I \wedge B \subseteq I$, while A is called prior and B is called posterior.

In our use case, items are all distinct functional genetic changes found in the library of available tumors. Item sets correspond to the set of functional changes of a single tumor and drug response classes obtained in Xenograft experiments. We are searching for rules $A \Rightarrow B$, where A is a set of functional changes and B is a specific drug response class. Our model focuses only on the impact of single functional changes to limit the problem space, i.e., we restrict $|A| = 1$.

We applied the Apriori algorithm for ARM using the Predictive Analysis Library (PAL), which is integrated in the IMDB, and the implementation provided by the R package `arules` [27, 28, 29].

For classification of drug responses, either the Tumor/Control (T/C) value or the Response Evaluation

Criteria in Solid Tumors (RECIST) value can be incorporated. In order to use Apriori ARM for classification, we defined the following drug response classes for each item set: Partial Response (PR), Stable Disease (SD), and Partial Disease (PD). The thresholds for the classes can be configured for each analysis individually by cancer researchers.

2) Classification of Tumor Data: Classification of tumor data can bring up hidden similarities in the data set, e.g., to generate hypotheses about new tumors subtypes. We provide the platform for execution of prediction models, but its concrete specification is provided by involved clinical researchers.

We use Support Vector Machine (SVM), which is available in many statistical frameworks, such as R [27]. It uses a regression mode also known as Support Vector Regression (SVR) to estimate correlation between attributes of the training data to re-apply them for prediction [30]. Our SVM implementation is part of PAL, which is executed directly in the IMDB and performs faster than existing alternatives due to incorporated technology advances, e.g., eliminated disk I/O and eliminating the need for exporting/importing data from/to the database.

In the following, we share details of our tumor classification. Firstly, the researcher configures SVM parameters through an interactive, web-based wizard, e.g., drug responses to predict or experiment data to use for training of the statistical model. The results depend on the configured parameters, e.g., a concrete T/C or a RECIST value for a specific pharmaceutical based on the selected tumor attributes. SVM in classification mode results in response class probabilities instead of concrete drug efficiency values as introduced in Sect. VI-C1 [31]. We define the response classes as follows:

$$0.0 \leqslant PR < 0.7 \leqslant SD < 1.2 \leqslant PD, \text{ with}$$

- **PR** defining a reduction in tumor growth,
- **SD** defining no significant change, and
- **PD** defining a negative drug response or a growth of the tumor.

The train formula in the R procedure is `"drug ~."`', which defines the `drug` attribute as the depending and the remaining database attributes as deciding variables indicated by the dot. Executing SVM for multiple drugs reuses the configuration to achieve a high level of parallel data processing.

VII. Evaluation and Discussion

Together with cancer researchers, we have been able to apply our enhanced research process. Major advantages are summarized in the following.

Integration of heterogeneous data sources: Instead of integrating latest data, e.g., genetic annotations, from international research databases manually, researchers can build on always up-to-date data maintained by the updater framework of the HIG platform [5].

Eliminated media breaks: By providing required statistical algorithms, e.g., clustering, as an integrated component of our IMDB platform, we reduced media breaks and required data transformations. Once raw data, e.g., experiment and genomic data, is stored in the IMDB, data processing is performed within the database. As a result, processing and analysis of data as the time most time-consuming operations are streamlined and the overall process time is significantly reduced.

Flexible and instantaneous data analysis: We build on the latest IMDB technology since it enables interactive analysis of big medical data due to its performance advances in data processing. Furthermore, it leverages flexibility in data analysis, i.e., user-configured parameters and aspects to analyze instead of having only access to a limited amount of optimized, but predefined and static analysis reports. Thus, interactive graphical data exploration supports researchers to identify and verify new hypotheses instantly without excessive delay.

Integrated statistical tools: We support the use of statistical tools directly within the IMDB. Thus, we were able together with cancer researchers to design and implement clinical models to predict Xenograft results based on obtained experiment data using SVM.

VIII. Conclusion and Outlook

In our contribution, we shared details about our enhanced research process for drug response analysis. We showed how latest IMDB technology acts as the key enabler for real-time data analysis, exploration of experiment data, and the integration of heterogeneous data sources. Thus, our HIG platform enabling processing and analysis of big medical data, is a foundation for implementation of the specific drug response analysis application while optimizing the existing research processes in this specific field of cancer research. Furthermore, we shared detailed insights in our applied research methodology, which constantly involves the feedback of experts from interdisciplinary teams.

Our future work will focus on applying the research process to additional fields of cancer research in course of precision medicine. Furthermore, we will investigate how a huge library of tumor samples can be used as training data to create more stable prediction models to discover new medical insights also for well-understood tumor types, such as breast or lung cancer.

References

[1] F. S. Collins *et al.*, “New Goals for the U.S. Human Genome Project,” *Science*, vol. 282, no. 5389, pp. 682–689, 1998.

[2] W. J. Ansorge, “Next-generation DNA Sequencing Techniques,” *New Biotechn*, vol. 25, no. 4, pp. 195–203, 2009.

[3] K. Jain, *Textbook of Pers Med.* Springer, 2009.

[4] M.-P. Schapranow *et al.*, “Mobile Real-time Analysis of Patient Data for Advanced Decision Support in Pers Med,” in *Proceedings of the 5th Int’l Conf on eHealth, Telemed, and Social Med*, 2013.

[5] M.-P. Schapranow, F. Häger, and H. Plattner, “High-Performance In-Memory Genome Project: A Platform for Integrated Real-Time Genome Data Analysis,” in *Proceedings of the 2nd Int’l Conf on Global Health Chall.* IARIA, Nov 2013, pp. 5–10.

[6] R. S. Kerbel, “Human Tumor Xenografts as Predictive Preclinical Models for Anticancer Drug Activity in Humans: Better than Commonly Perceived But They Can Be Improved,” *Cancer Biology & Therapy*, vol. 2, no. 4, pp. 134–139, 2003.

[7] National Human Genome Res Inst, “DNA Sequencing Costs,” http://www.genome.gov/sequencingcosts/[2], Apr 2013.

[8] J. C. McCallum, “Memory Prices (1957-2014),” http://www.jcmit.com/memoryprice.htm[2], Apr 2014.

[9] Y. Sun, “Identification of REST Regulated Genes in Prostate Cancer via High-throughput Technologies,” Master’s thesis, Instituto Superior Técnico, Universidade Técnica, Lisboa, Portugal, Dec 2012.

[10] F. J. Rossello *et al.*, “Next-Generation Sequence Analysis of Cancer Xenograft Models,” *PLoS ONE*, vol. 8, no. 9, p. e74432, Sep 2013.

[11] H. Li and R. Durbin, “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transformation,” *Bioinform*, vol. 25, pp. 1754–1760, 2009.

[12] A. McKenna *et al.*, “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data,” *Genome Research*, vol. 20, no. 9, pp. 1297–1303, 2010.

[13] P. Cingolani *et al.*, “A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms,” *Fly*, vol. 6, no. 2, pp. 80–92, 2012.

[14] H. Plattner, *A Course in In-Memory Data Management: The Inner Mechanics of In-Memory Databases*, 1st ed. Springer, 2013.

[15] P. Svensson, “The Evolution of Vertical Database Architectures: A Historical Review,” in *Proceedings of the 20th Int’l Conf on Scientific and Stat Database Mgmt.* Springer-Verlag, 2008, pp. 3–5.

[16] S. S. Lightstone, T. J. Teorey, and T. Nadeau, *Physical Database Design: The Database Professional’s Guide to Exploiting Indexes, Views, Storage, and more.* Morgan Kaufmann, 2007.

[17] J. M. Hellerstein, M. Stonebraker, and J. Hamilton, *Architecture of a Database System, Foundation and Trends in Databases.* now Publishers, 2007, vol. 1.

[18] A. R. Hevner *et al.*, “Design Science in Inform Syst Res,” *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.

[19] H. Plattner *et al.*, *Design Thinking Research*, ser. Understanding Innovation. Springer, 2012.

[20] R. Pichler, *Agile Product Mgmt With Scrum: Creating Products that Customers Love*, ser. Agile Software Development Series. Addison-Wesley, 2010.

[21] S. Oesterreich, A. M. Brufsky, and N. E. Davidson, “Using Mice to Treat (Wo)men: Mining Genetic Changes in Patient Xenografts to Attack Breast Cancer,” *Cell Rep*, vol. 4, pp. 1061–1062, Sep 2013.

[22] N. Rani, R. Singh, and G. Arora, “Detection of ORF Frames Using Data Mining,” *Int’l Journal of Computer Science and Telecommun*, vol. 2, no. 3, pp. 90–94, Sep 2011.

[23] SAP AG, “SAP HANA SQL and System Views Reference SPS08 Ver. 1.1,” http://help.sap.com/hana/SAP_HANA_SQL_and_System_Views_Reference_en.pdf[2], Aug 2014.

[24] A. Hannan, “The L Programming Language & System,” http://home.cc.gatech.edu/tony/uploads/61/Lpaper.htm[2], Jan 2005.

[25] C. Bresch and R. Hausmann, *Klassische und molek. Genetik*, 3rd, Ed. Springer-Verlag, 1972.

[26] F. Hsu *et al.*, “The UCSC Known Genes,” *Bioinformatics*, vol. 22, pp. 1036–1046, 2006.

[32] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data Clustering: A Review,” *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[27] SAP AG, “SAP HANA R Integration Guide SPS08 Ver. 1.1,” https://help.sap.com/hana/SAP_HANA_R_Integration_Guide_en.pdf[2], Aug. 2014.

[28] ——, “SAP HANA Predictive Analytics Library (PAL) SPS08 Ver. 1.1,” https://help.sap.com/hana/SAP_HANA_Predictive_Analysis_Library_PAL_en.pdf[2], Aug. 2014.

[29] M. Hahsler, B. Grün, and K. Hornik, “Introduction to arules: Mining Association Rules and Frequent Item Sets,” in *Special Interest Group on Knowledge Discovery and Data Mining*, 2007.

[30] A. J. Smola and B. Schölkopf, “A tutorial on Support Vector Regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[31] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[2] All online references were checked on Sep 8, 2014.