

Design implementation of MAV-clic: Architecture, Components and Build

Supplementary Material: “MA V-clic: Management, Analysis and Visualization of Clinical Data”

Authors List:

Zeeshan Ahmed, PhD¹,

¹Department of Genetics and Genome Sciences, School of Medicine, University of Connecticut Health Center, Farmington, CT, USA;

Minjung Kim, PhD²

²The Pat and Jim Calhoun Cardiology Center, School of Medicine, University of Connecticut Health Center, Farmington, CT, USA.

Bruce T. Liang, M.D., F.A.C.C.³

³Ray Neag Distinguished Professor of Cardiovascular Biology and Medicine, School of Medicine, University of Connecticut Health Center, Farmington, CT, USA.

Corresponding Authors:

Zeeshan Ahmed, PhD,

Assistant Professor & Assistant Director Bioinformatics Precision Medicine Program,
Department of Genetics and Genome Sciences, School of Medicine, University of
Connecticut Health Center, 400 Farmington Ave, Farmington, CT, 06032, USA.

Email: zahmed@uchc.edu

Phone: 860-679-2643

29

Design implementation of MAV-clic: Architecture, Components and Build

Zeeshan Ahmed, PhD¹; Minjung Kim, PhD²; Bruce T. Liang, M.D., F.A.C.C.³

¹Department of Genetics and Genome Sciences, School of Medicine, University of Connecticut Health Center, Farmington, CT, USA; ²The Pat and Jim Calhoun Cardiology Center, School of Medicine, University of Connecticut Health Center, Farmington, CT, USA. ³Ray Neag Distinguished Professor of Cardiovascular Biology and Medicine, School of Medicine, University of Connecticut Health Center, Farmington, CT, USA.

Abstract

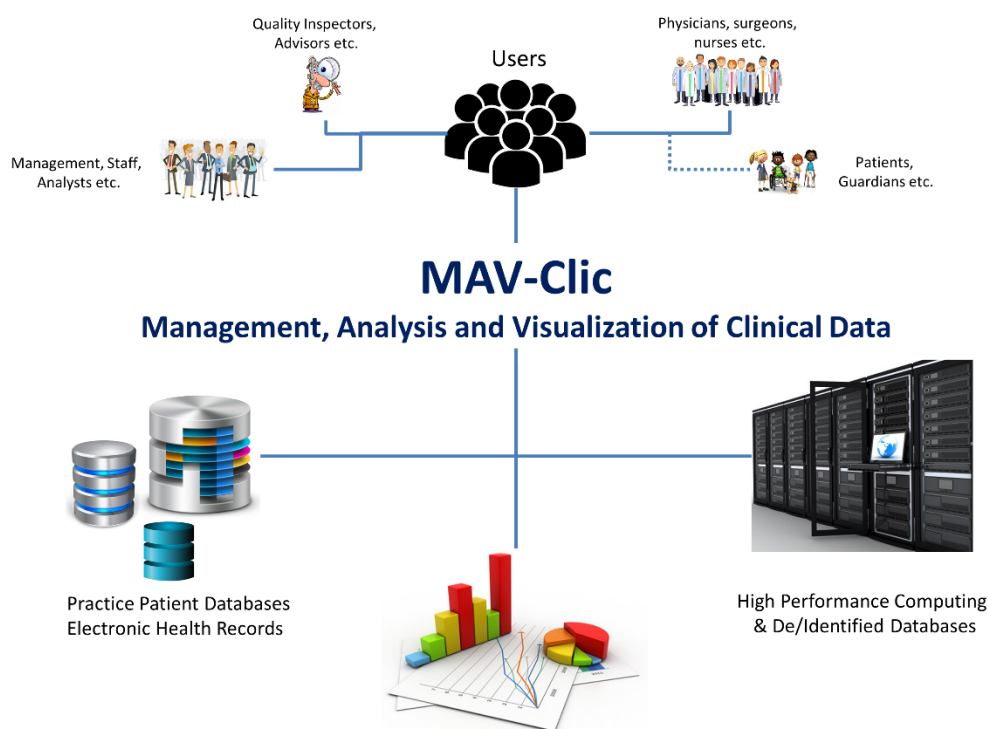
Healthcare data includes information about patient life style, medical history, visits to the practice, lab tests, imaging test, diagnoses, medications, surgical procedures, genomics profile, consulted providers and claims. Adequate and analytic access to the health care data has potential to revolutionize the field of medicine by improving the quality and transition of care, and developing better understanding of biological mechanisms and modelling complex biological interactions by integrating and analyzing knowledge in a holistic manner. To fulfill the growing interests in implementing the health information system, MAV-clic is developed to satisfy the requirements of data owners as well as data users in the healthcare system. As the multi-database management system, it can benefit for the data owners to manage huge database from multiple sources in a centralized manner. The data having different contents, formats, styles, sizes, structures can be extracted and transformed into the normalized format and stored into the MAV-clic system with High performance computing technology. The well-organized data management features allow for data users to analyze the complex, disparate healthcare data. Analytics process in MAV-clic can help building the cohort in terms of demographic information, and time information as well as generating evidence using analytics of the patient's information along with diagnoses, medications, laboratory results. MAV-clic also offers the customized functions, which can explore the quality measures of the hospital using the EHR database, visualize the patterns of the analyzed results, and report the summary in an automated and timely manner. In this manuscript we present the concept, system design, modelled database, data flow, components, and software development, deployment and configurations details of MAV-clic.

Keywords:

Analysis, Database, Data Mining, HIPAA, Patient Generated Health Data

MAV-clic Concept

The overall concept of MAV-clic is to help support and implementation of new healthcare data analytic and research process at the UConn Health. The goal is to produce a prototype application and database management system, which can connect people from different background and specialists (e.g., *Podiatrists, Optometric, Oral Surgeons, Dentists, Chiropractors, Physician Assistants, Nurse Practitioners, Clinical Nurse Specialists, Clinical Social Workers, Clinical Phycologists, Registered Dietitians, Nutrition Professionals Physical Therapists, Occupational Therapists, Qualified Speech-Language Therapists, Quality Inspectors, Management, and other staff members* etc.) with direct access to the electronic health records in a way that they ask some analytic questions, which can be helpful in drawing those pictures, which a traditional electronic health records system cannot easily facilitate.

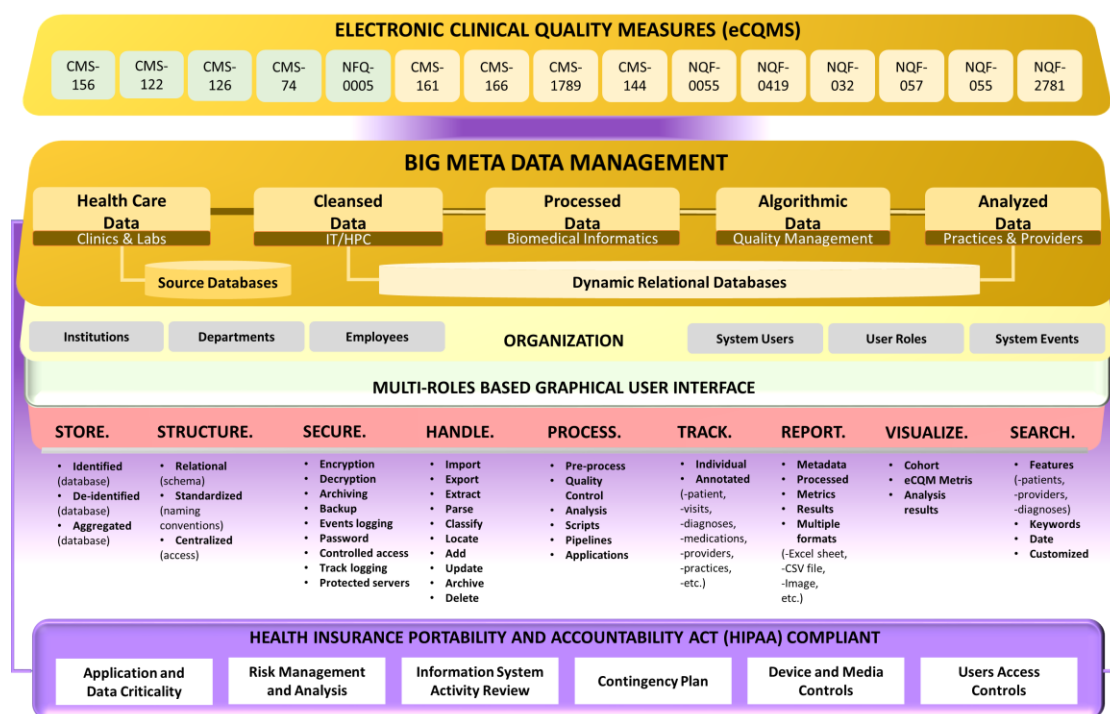


S. Figure 1. Concept of MAV-clic

As presented (S. Figure 1), MAV-clic provides multi roles based users access to the system, facilitates electronic health records transfer using high performance computing environment and allows users to ask simple and complex questions to the system and report results.

MAV-clic System Designs

MAV-clic is a new, large scaled, user-friendly, interactive, cross-platform, encrypted, multi-roles based, automated, customized and centralized multi-database management system at the UConn Health. MAV-clic is based on Butterfly Model^{1, 2} and product line architecture³ (S. Figure 2). It implies that all major modules are capable of performing individual key roles as well as integrating each other.



S. Figure 2. Architecture of MAV-clc

While design and implementation of the system, one of the most complicated tasks were to implement Health Insurance Portability and Accountability Act (HIPAA)^{4, 5, 6, 7} compliant environment, as it involves not less than a thesis to be pre-compiled with all instructions, and guidelines strictly followed and to be implemented in the system following HIPAA rules. MAV-clc is a HIPAA compliant framework that implements healthcare and users data security, which includes: *Application and Data Criticality; Risk Management and Analysis; Information System Activity Review; Contingency Plan; Device and Media Controls; and Access Controls.*

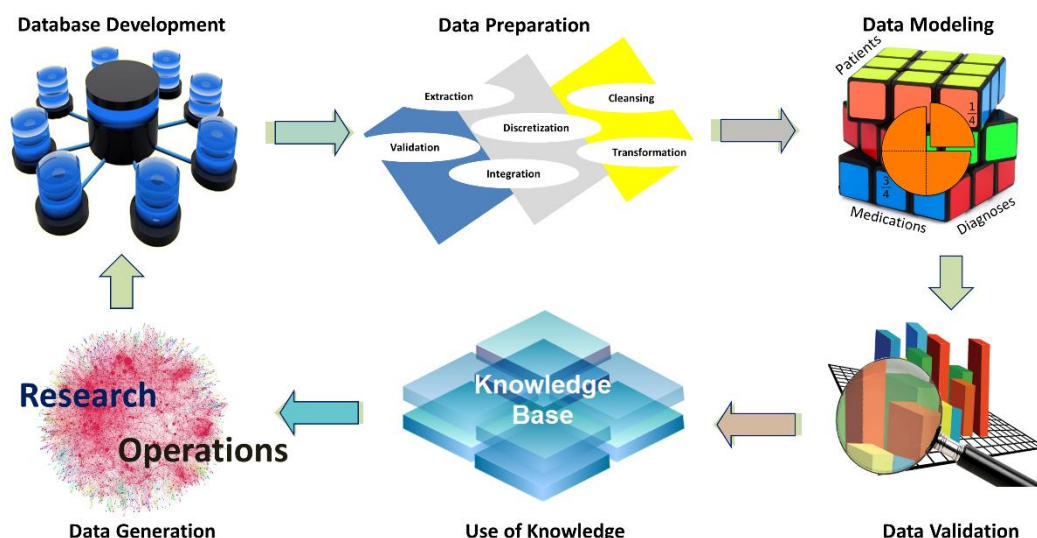
User must be one of the trusted ones but that is not enough. We have implemented very tight security to ensure data privacy, which requires many conditions e.g., user must be cleared from the IT/Security departments to have access to the healthcare data, user can only use MAV-clc when installed within a computer located within HIPAA compliant environment (only intranet based, no WIFI or internet connection works) etc. Moreover, we have implemented strong security checks within system to monitor user's activities and to manage alumni users and obsolete system's versions. MAV-clc implements well monitored security around system. We have provided a special graphical interface to manage users' information, which only admin can access to create, restrict, block, fully track and update users' account information.

The system overhauls the status of version automatically. New features are added in the system and it is required to make sure that all users have correct and updated version. Considering the fact that software can be used by users who are scattered throughout the UConn Health campus, it's very hard to know which user has the right version to use. We have added features in the system to make sure that all users have correct and updated version. System will automatically restrict and direct him to the latest version, when a user will try to login with old version.

Moreover, system generated data is fully encrypted, even administrator cannot see the confidential information. Users can download system generated data in a de-identified manner while users have read-only access to the identified data. System generated data is fully encrypted so even administrator cannot see the contents. Also, system generated data is composed of confidential information, so it can only be archived and retrieved back but cannot be permanently deleted from the system. We have enforced some security conditions, which includes:

- 1) All the events fired and operations performed by the users are fully logged, even all the system generated SQL queries are recorded.
- 2) Users are not allowed to download patients' personal details.
- 3) There is a limit applied in even downloading data.
- 4) Users can only access desktop application only when they are connected using Ethernet are in secure and permitted locations/labs.
- 5) The system is audited at pre-determined time intervals. So, all the users are followed and tracked for their operations periodically.

MAV-clc implements healthcare data analytics process (S. Figure 3), which includes review and finalization of study documents, patient database designing, data collection, data preparation, data tracking, data modelling, data validation, discrepancy management (process to systematically correct and manage discrepancies generated within data to ensure that data is complete, accurate, and compliant), medical coding, knowledge base, quality metrics management, database locking and users' access control and management. Data is raw and once it's extracted, information is collection for analysis and once its analyzed and meaningful results are available and stored in the system, which are the part of knowledge base e.g., reports including information about customized eQCMs, denominators, numerators and percentages etc.

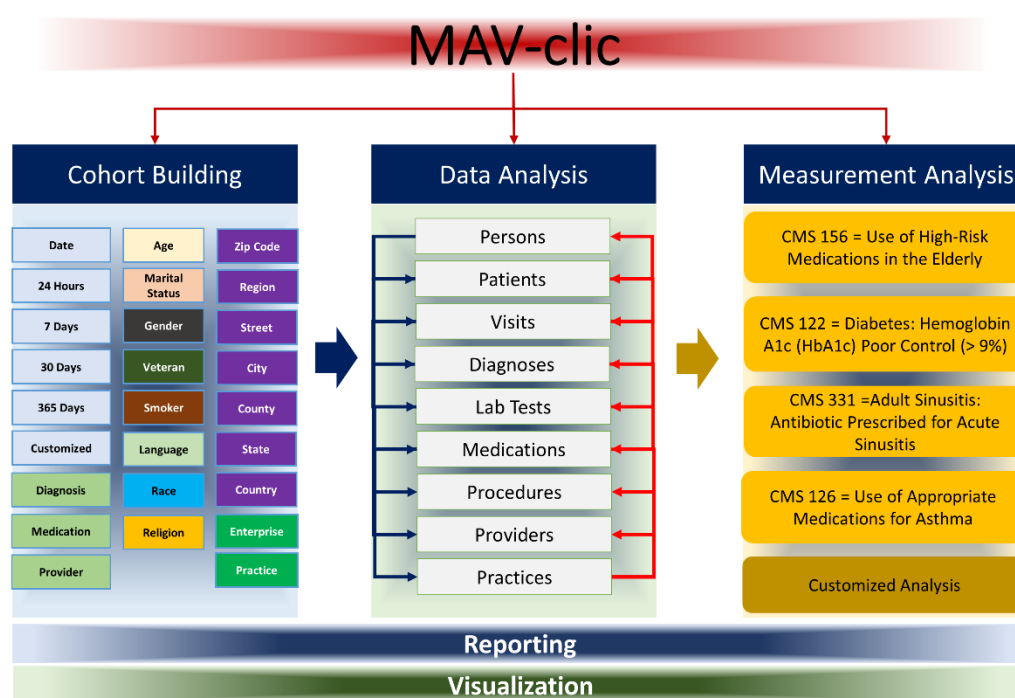


S. Figure 3. Health care data analytics process

The major analytics components dealing with the clinical data in MAV-clic consist of three different modules: *Cohort building*, *Data analysis*, and *Measurement analysis* (S. Figure. 3)

In cohort building module, multiple demographic features are offered to build cohort and ontology from registered patient population. Cohort can be made based on the days to visit the practices, personal information (e.g., gender, age, marital status, language, race, religion, smoking and other habits), regional information (zip code, street, city, county, state, and country), diagnoses, medications, procedures and consulted providers etc. Once the cohort is built, care providers can analyze their patients data based on the diagnoses codes, lab test results and prescribed medications in data analysis module. It is capable of tracking one specific patient in the selected cohort as well as the cohort itself in terms of visit to the practices, diagnoses, medication, and consulted providers.

In measurement analysis module, customized functions are offered to search for the patients by the lab test results (the name of lab test, the range of lab test results), or prescribed medication information (the name of medication group, the number of medication, medication duration). The sorted out results based on the selected criteria are counted and used to calculate the quality measures. Then, the aggregated results was reported in a de-identified form. Thus, healthcare organizations can report the electronic clinical quality measures (eCQMs) that the Center for Medicare and Medicaid Services (CMS) has proposed in order to help unlocking the insights from patient data, demonstrating values in evidence-based medicine, improving outcomes and reducing costs (<https://ecqi.healthit.gov/ecqms>).

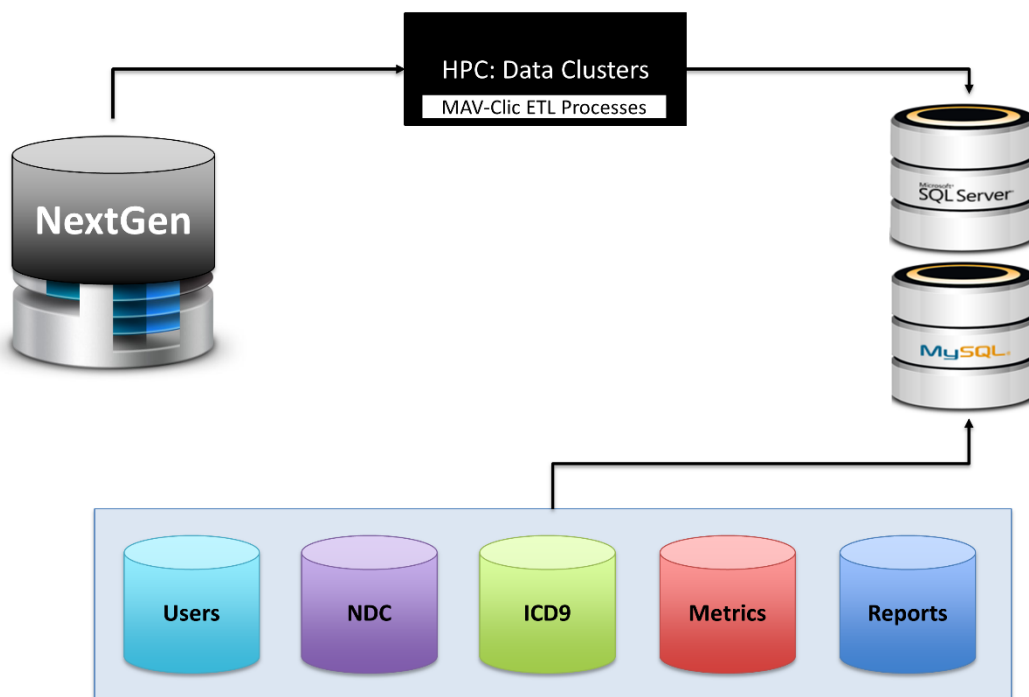


S. Figure 4. Components of MAV-clic

MAV-clic Database Components

Commercial electronic health records (EHR) systems (e.g., *NextGen*, *Epic*, *Cerner*, *GE Healthcare*, *eClinicalWorks*, *athenahealth*, *McKesson*, *Allscripts*, *Care360* etc.) are very complicated, closed and restricted. These are designed for commercial purposes with goals mainly circulating around the data entry and retrieval without comprehensive analytics implementation. In our case, MAV-clic efficiently extracts data from NextGen having over 7000 relations and other in-house databases, then cleanses and stores the extracted data in a format of Microsoft SQL, MySQL, Data clusters.

MAV-clic implements multiple and different databases which have different roles in managing the data and link them together to function as one integrated system (S. Figure 5). The current version implements Microsoft SQL server for EHR data management and MySQL server for users, national drug codes (NDC), international classification of disease, ninth revision Codes (ICD9), metrics and reports management.



S. Figure 5. Databases of MAV-clic

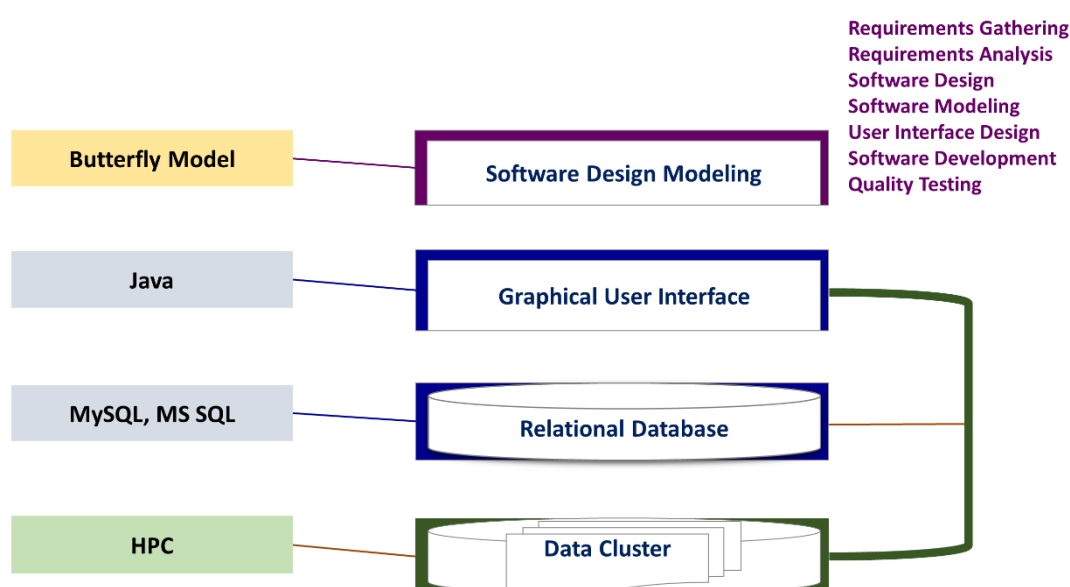
MAV-clic implements ETL processes, which are capable of running in command line HPC environments as well as in the desktop computers to extract, transform, and load the source of clinical data to the destinations of MAV-clic servers. During ETL processes, the data is structured (normalized relational schemas, standardized naming conventions) and secured in term of the encryption, archiving, backup, event logging, limited access and password management in identified and de-identified formats. Furthermore, it is able to handle data with the basic functions (export, import, extract, parse, classify, search, add, update, archive, delete), track the data in an individualized and annotated manner, and visualize and report the analyzed results with common formats such as Microsoft Excel spreadsheet, common separated values (CSV), text, images, PDF, html, xml files etc.

MAV-clic Software Development

MAV-clic is based on Butterfly Model^{1, 2} and product line architecture³. It is completely developed using JAVA programming language. MySQL and Microsoft SQL database management systems are used for the implementation of database (S. Figure 6). MAV-clic is tested, in use and recommended for Microsoft Windows and Macintosh operating systems.

Installation and Configuration

MAV-clic is programmed in Java programming language, built and tested at both Mac-OS-X and Windows platform. To successfully run (JAR file), it requires Java Version 8 with Update 131 (build 1.8.0_131-b11) can be downloaded from www.java.com. Additional information and updates are available at: <https://github.com/drzeeshanahmed/MAV-clic/wiki>. For further assistance, it is advised to contact authors: MK and ZA.



S. Figure 6. MAV-clic development tools and technologies.

References

1. Ahmed Z, Zeeshan S, Dandekar T. Developing sustainable software solutions for bioinformatics by the “Butterfly” paradigm. *F1000Research* 2014;7:54-66.
2. Ahmed Z, Zeeshan S. Cultivating Software Solutions Development in the Scientific Academia. *Recent Patents on Computer Science* 2014; 7:54-66.
3. Ahmed Z. Designing Flexible GUI to Increase the Acceptance Rate of Product Data Management Systems in Industry. *International Journal of Computer Science & Emerging Technologies* 2011; 2:100-109.
4. Turner S, Foong S. Navigating the Road to Implementation of the Health Insurance Portability and Accountability Act. *American Journal of Public Health* 2003;93:1806–1808
5. Miller JD. Sharing clinical research data in the United States under the health insurance portability and accountability act and the privacy rule. *Trials* 2010;11:112.
6. Goldstein MM. Health Information Privacy and Health Information Technology in the US Correctional Setting. *American Journal of Public Health* 2014;104:803–809.
7. Bradford W, et al. Development of a HIPAA-Compliant Environment for Translational Research Data and Analytics. *J Am Med Inform Assoc* 2014;21:185–189.

227 **Acknowledgements**

228 We are grateful to the University of Connecticut Health Center (UConn Health), School of
229 Medicine, Department of Genetics and Genome Sciences, Institute for Systems Genomics, The
230 Pat and Jim Calhoun Cardiology Center, and Cardiovascular Biology and Medicine, SNE-PTN
231 and Ahmed Lab. We appreciate all colleagues, who provided insight and expertise that greatly
232 assisted the research and development.

233

234 **Authors and contributions**

235 ZA perceived the idea and did all work on the software and infrastructure design and
236 implementation and related aspects of MAV-clic. ZA and MK did analysis and performance
237 evaluation of MAV-clic. BL guided study.