Cloud MRM: A bio-IT tool for correlative studies in dermatology

Yan Valle, M.Sc, MBA
Prem Couture, M.Sc
Torello Lotti, Professor, PhD, MD
Igor Korobko, PhD, D.Sci

**Corresponding author. E-mail: yan@vrfoundation.org*

Abstract
Innovative medical information solutions include, among others, Personally Controlled Health Records, which enable a patient to manage a secure copy of his or her medical data, and Cloud Computing, which provides secure storage and quick analysis of large datasets. Here we present Cloud MRM bio-IT system for secure storage and cross-analysis of granulated patient data and large genomic datasets. We describe how Cloud MRM will help to uncover correlations between molecular parameters such as single nucleotide polymorphism profile and clinical diagnosis in vitiligo, while enabling registration of all relevant medical data in support of downstream therapy development. In the recent release version 1.0 we have included a vitiligo-specific electronic health record, a database of selected genes and their polymorphisms, and data analysis tools. Thus a transparent R&D environment is built and tested to encourage widespread collaboration of patients, researchers, and healthcare providers for expedited development of innovative, patient-focused therapies as well as to facilitate understanding of pathogenic mechanisms of vitiligo.

Keywords: Vitiligo, leukoderma, bioinformatics, cloud computing, bio-IT.

Introduction

Vitiligo is one of the most common pigmentary disorders affecting at average 0.5-2%, and up to 8% of population as recently reported in some regions, such as India. While mostly considered as cosmetology problem, research in this field demonstrated that this is a systemic disease.

Non-segmental vitiligo (NSV) is the most common type of disease which is acknowledged to be driven by complex and currently unknown pathogenic mechanisms. Data accumulated by now undoubtedly show that despite clinically similar manifestation, molecular pathogenesis of vitiligo is far from being uniform and different molecular mechanisms are likely to underlie the disease onset and progression. Yet current treatment strategy selection for NSV is not based on NSV molecular features of the particular patient. This challenges for research aiming to uncover specific molecular features of vitiligo pathogenesis which would enable personalized approach in vitiligo treatment and facilitate new target identification for rational drug discovery.

Completing this task would definitely require collection of large datasets of different types of data (clinical, biochemical, histological, genetic, etc.) and their analysis aiming to reveal common molecular abnormalities for subgroups of patients

and their association with clinical manifestation of the disease and particular treatment efficiency. Accomplishing this task requires large variegated datasets collection, storage and analysis, mostly to identify correlative links.

A recently updated Cochrane review [24] reports only total of 57 studies on vitiligo in recent years, with 3139 participants overall. Most of the studies had fewer than 50 participants and few lasted longer than 6 months. Only 28 of the trials reported quality of life and greater than 75% repigmentation. Only two of the 57 analyzed studies could be combined for meta-analysis, as usefulness of the findings is limited by different designs and outcome measurements and lack of quality of life measures.

In the case of vitiligo, R&D process is too time-consuming due to the lengthy disease course and too costly due to the high number of required patients. Even more critical for the development of vitiligo treatment is the ability to test the complex combinations of therapies responsible for the substantial improvement in long-term outcome, which is now essential to the vitiligo treatment process.

We are now poised for the big transformation of pharmaceutical industry, which will come from behavioral changes, as all participants in the ecosystem - patients, physicians, payers, companies and more - revisit and realign their practices in order to improve health outcomes.

Today, many innovative medical information solutions emerge [1] and are put into practice, such as: Personal Health Record Systems, Health Information Exchange Networks, Genomic Information Systems & BioRepositories, Wearable and Implantable Health IT Systems, Cloud Computing. These innovative solutions have great potential to improve quality of care, expedite research, lead to faster drug discovery, increase patient safety, and lower costs of medical R&D over the long term for various disease conditions.

We have identified at least 14 disease categories in which companies have launched new consumer-facing applications in the last year alone. These include apps that help patients keep track of vaccination schedules (Novartis' VaxTrak), manage their hemophilia A Factor VIII infusions (Bayer's Factor Track), locate cancer clinical trials within 150 miles (GSK's Cancer Trials) and much more. SanofiAventis launched AFib Educator, an iPhone app that helps health care providers explain atrial fibrillation to patients, their families and caregivers. And Japan's Astellas has released a smartphone app that gives physicians access to criteria used to assess the need for cardiac radionuclide imaging.

Genetic data are becoming a routine component of clinical diagnosis and treatment in ever increasing number of diseases [3]. Recent research [4] demonstrated that several treatment responses are conditional on genomic profile, and promises to usher in the long-awaited era of personalized medicine, all based on the patient's gene sequence or gene expression signature. Thus these types of data are wisely to be included in analyzed datasets to approach vitiligo problem. The cost of genome expression profiling and genotyping is falling rapidly with high-throughput techniques [5, 6] and becoming affordable for individuals. For example, individual genome sequencing cost was recently (as of June 9, 2011) knocked down from $48,000 to $7,500 per genome [11]. There are even genotyping cost calculators available online [7]. It is likely that gene expression levels will be measured for many serious ailments in the next coming years. The tremendously increased dataflow

which has to be taken into account in data analysis demands innovative solutions in bio-IT to conform to arising requirements.

The impact of cloud computing [15, 16] is just beginning to be felt in the areas of research, development, clinical trial management and healthcare information exchanges. The explosion of data from next generation sequencing, the growing importance of biologics in the research process is making cloud-based computing an increasingly important aspect of R&D. For example, complex genetic sequences and biomarker data are now being hosted in the private or public cloud, such as Amazon Simple Storage Service, or Microsoft Azure Clouds [17]. Data are then accessed in a secure fashion by companies for their research needs.

Here we present the design and implementation of the Cloud Medical Research and Management (Cloud MRM) system and demonstrate how the system can be used as a standard tool for vitiligo R&D, and become a vehicle for scientific collaboration and decision support.

Results and Discussion

Cloud MRM system overview

Cloud MRM is primarily designed as tool for researchers to uncover correlations between gene expression patterns and clinical diagnosis, such as vitiligo, while enabling registration of all relevant medical data in support of downstream therapy development. It may also accommodate "information altruists" [18] who are willing to share some of their personal, clinical and genomic characteristics for expediting vitiligo research.

Our most recent Cloud MRM deployments are focused on providing the cloud-based platform for patient-centered integration and analysis of health information related to vitiligo, and building a collaborative environment for researchers, to improve the quality, effectiveness, and convenience of dermatological R&D. Currently, Cloud MRM provides integration of: (i) detailed clinical records, (ii) laboratory test results, (iii) database of investigated genes and polymorphisms, (iv) online survey with branching logic, (iv) Deep Zoom photo tool, (v) statistical analysis tools, for cross-analysis of phenotype, genotype and blood test results, which may help to uncover unseen relationships in vitiligo. The Cloud MRM is designed to be interoperable with electronic medical records and other information systems by using widely-adopted, standardized methods for exchanging data, such as LIONS for color coding, CCR and CCD for information transfer, HITSP for interoperability.

On the front-end, Cloud MRM is a web-based Personal Health Record System (Fig. 1). It allows patients, researchers and healthcare providers to input, share and cross-analyze health record data. On the back-end, Cloud MRM is a low-level storage and high-level medical document organization, in order to enable the secure storage and statistical analysis of fine-grained patient data (Fig. 2). Non-personalized vitiligo electronic health record is built-in and available to all users by default.

Fig. 1. On the front-end, Cloud MRM is a web-based Personal Health Record System.

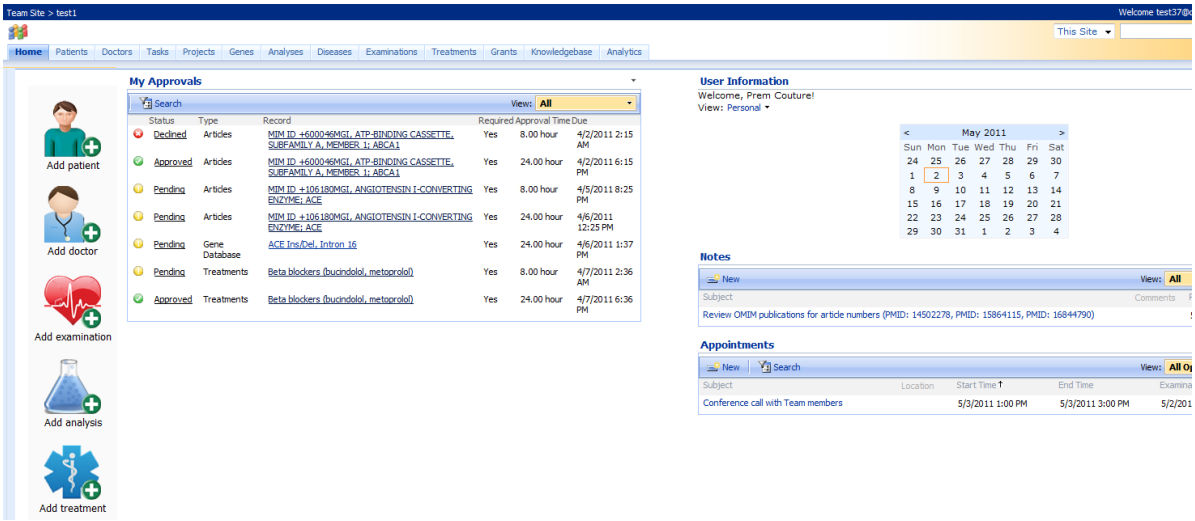

Fig.2. On the back-end, Cloud MRM is a low-level storage and high-level medical document organization.

The reference implementation of Cloud MRM provides a simple interface for patients and healthcare providers using the internet. User interfaces can be highly customized in look and feel and provide targeted views of health data for various subsets of the vitiligo community. For example, user may only see a portion of his or her record, or distinctive workflows with specialized data type management may be implemented for R&D personnel. The highly granular design of the vitiligo health record greatly facilitates analysis of de-identified records by participating researchers.

Data

Our pool of data is designed to be derived from the across the medical ecosystem through diverse types and sources of data, such as various health-record platforms, social-media conversations, wireless-device biometric information, etc. (Fig. 3).

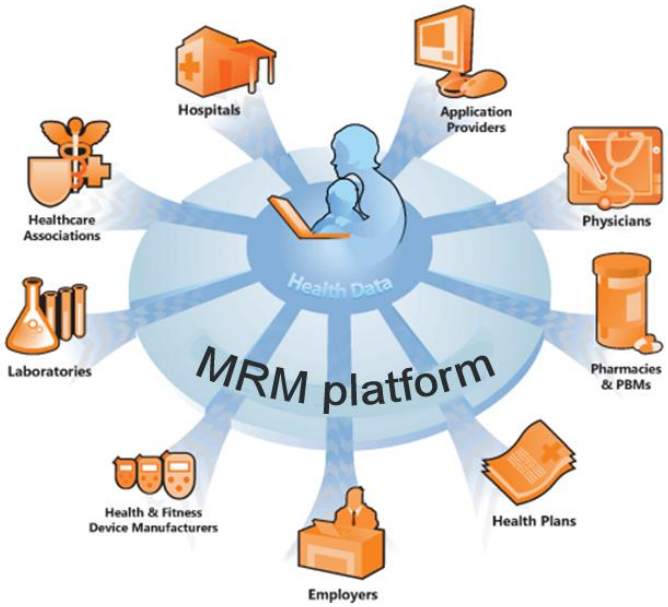

Fig.3. Our pool of data is designed to be derived from the across the medical ecosystem.

Currently, a document within the Cloud MRM represents: personal profile with 35 individual values, medical profile with 240 values, laboratory blood test result with 150 values and deviation norms, survey result with 210 values, database of 250 selected genes and investigated polymorphisms, high resolution images in jpeg format, clinical encounters and medications. It may include any other discrete piece of relevant healthcare information, which we will decide to encode into the system.

In the Cloud MRM model, user record is populated as the patient provides answers to online questionnaire with branching logic, fills out forms, uploads images, or as lab returns test results back into user record.

Since the Cloud MRM is institutionally neutral, it encodes laboratory test result values in their original format and provides conversion into internationally recognized standards upon user request. Custom deviation from recognized standard values can be input manually by researchers, allowing personalized research strategies.

Researchers using the Cloud MRM system must be confident in an accurate and trustworthy pool of data, anonymously derived from the patient's health history. Hence, we limit content modification. The system will not allow the patient to modify a lab test value returned by a registered healthcare provider. The lab test results provided by the user will bear internal “Unconfirmed” status until verified and confirmed by the healthcare provider or system administrator.

Genomic data is treated as a very large sequence of small results. The most efficient way to genotype "most of" a human being is to genotype tag SNPs according to the HapMap [20], which is expected to highlight close to 600,000 SNPs that identify most of the clinically-useful genetic diversity. Considering necessary redundancy for indexing of partial records, the size of a patient's genome sequence might be in proximity of 5 megabytes.

Users, data access and deposition

To conform to privacy policy, a numerical identificator is assigned to each patient by the system administrator or physician, and only those two are aware of personality hidden under identificator. Other users of the system are able to use de-identified extracts of data from patient records without explicit individual consent. If desired, patient can temporarily disclose his name in the record and can indicate which other users have particular privileges on specific portions of their record. This policy affords maximum control by the record owners over their personal health data within the boundaries set by the system administrators.

To create an account, we implemented an institution-based policy that only allows system administrator to create new accounts. Each account includes three types of information fields: personal details, clinical details and results of laboratory tests. Personal details in the account are filled by the patient or by his physician, and this is the only category of account data that can be edited by the patient with the read-only access to other data. All other types of data are entered by the qualified contributors within their competence as restricted by the administrator. For example, physicians can edit clinical data while authorized laboratories can enter data of analysis. Researchers not contributing to data collection may only aggregate de-identified data to search for patterns and complex relationships.

Data analysis

The Cloud MRM is designed to be able to quantitatively estimate the complex interplay of severity, progression, and other clinical factors in relation to treatments. The relationship between a blood test and a clinical outcome will rarely be a perfect linear or proportional. But we cannot and should not expect the system to accurately and quantitatively predict all clinical outcomes of our data analysis.

For analysis we use data from comprehensive vitiligo patient profile, including some surrogate endpoints. These endpoints are directly related to the disease pathophysiology and also to the mechanism of therapy action, and more accurately assess a beneficial treatment effect. The surrogate can provide a clear indication that the treatment effect is occurring and that the probabilistic outcome of clinical benefit is present (Fig.4a,b.)

We are not suggesting methods to identify or qualify surrogate endpoints. Some uncertainty for surrogates will always exist in rare diseases with limited patient-based clinical data. By highlighting and quantifying results generated by the Cloud MRM, we hope to incentive experts in this field to discuss practical and rational solutions to qualifying surrogates as primary endpoints for pivotal shift in vitiligo R&D.

Fig.4a. Dashboard screenshot of the analysis tools page – on PC and iPad screens. The surrogate can provide a clear indication that the treatment effect is occurring and that the probabilistic outcome of clinical benefit is present.

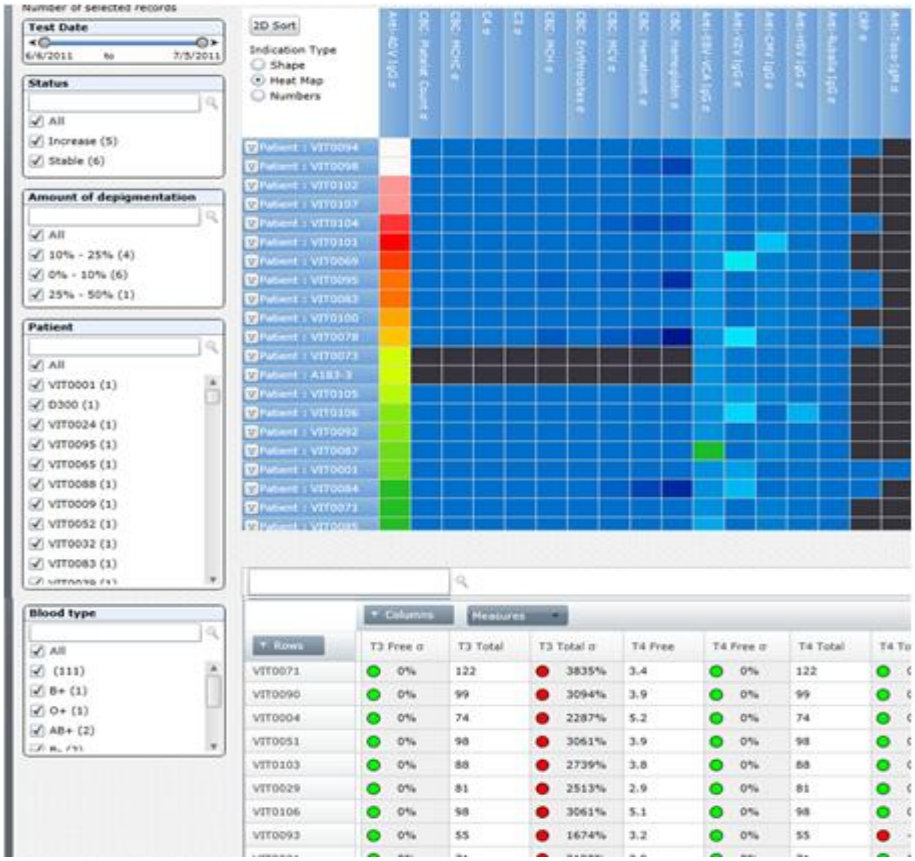

Fig. 4b. Detailed screenshot view of the "heat map" analysis tool and measured deviation from the median parameters.

Advanced graphical interface enables users to analyze patient profiles and genotypes to discover relationships and patterns, i.e. to stratify patients by blood test results or genotype that co-relate to disease progress (Fig.5.)

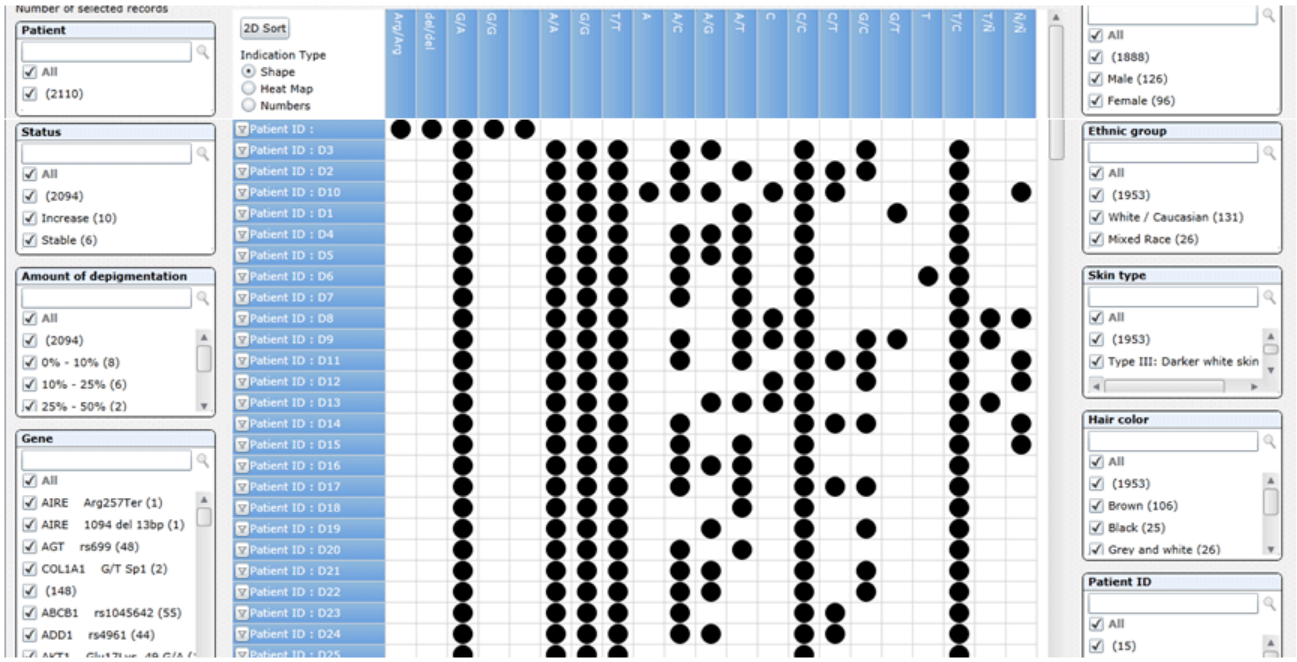

Fig. 5. Patients stratification by genotype results

In addition, Deep Zoom Imagery (Fig. 6) analyzes and compares progression of disease by patient, genotype, treatment, stages and other criteria. High Resolution (100MB+ per image) images converted to Deep Zoom format for fast vizualisation and analysis of over 10,000 images in 1-2 seconds.

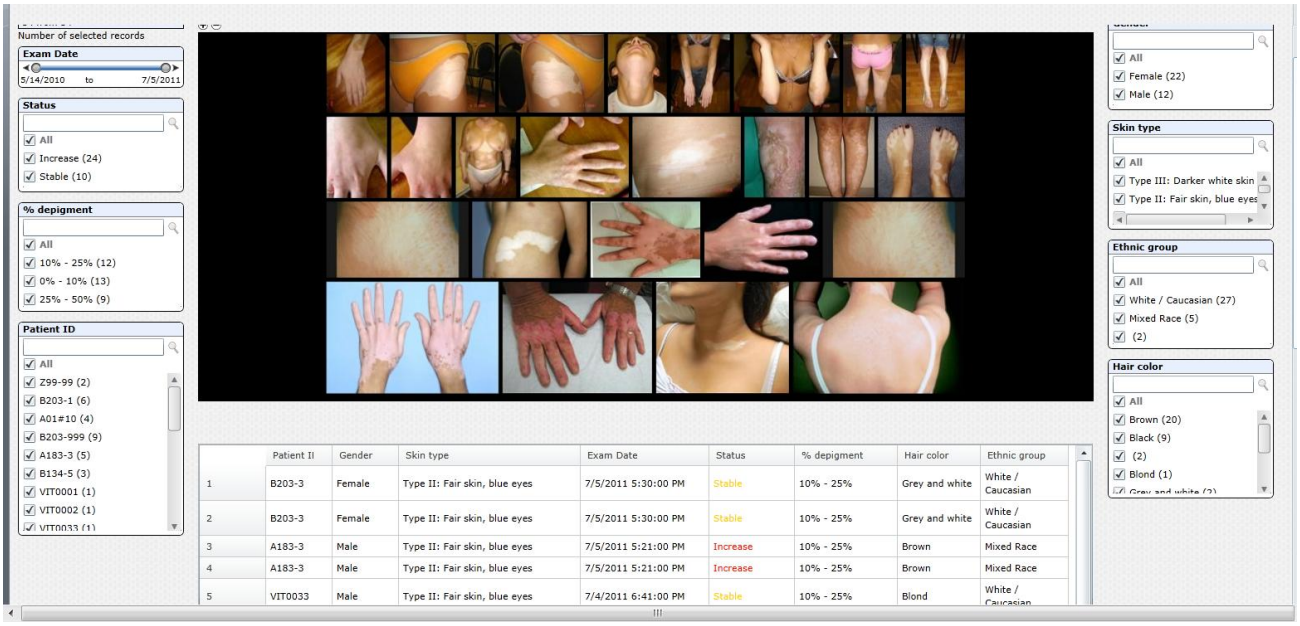

Fig. 6. Deep Zoom analyzes and compares progression of disease by patient, genotype, treatment, stages and other criteria.

Security

As healthcare ecosystem adopts new technologies, it also faces new security threats. Hackers, computer viruses, disaffected employees, and human error present real dangers to healthcare networks.

It is impossible to build a 100% secure computer system. A threat model was defined to analyze the Cloud MRM privacy requirements and ensure that the level of security provided by our implementation is reasonable. It was demonstrated [22] that there is a critical tradeoff between tight control and accessibility of the record for medical care. In the Cloud MRM, physical data store level itself persists as an encrypted object, thereby reducing the threat of disclosure from discarded or mismanaged disk drives that have been widely reported [23]. All data is secured with straight-forward AES encryption [21] to protect it against unauthorized access in the event of external attack, hardware mismanagement or data theft by employees.

We minimize this risk by storing in the principal database de-identified patients records only, while keeping the encrypted database with personal information in a separate secure environment.

Our strategy results in a layered approach to security, in which the failure of one system is not likely to lead to the compromise of network resources. Our secure network architecture includes protection for the network perimeter, local office network, teleworker connections, and satellite or remote locations.

Further development of the system

The recent release of Cloud MRM version 1.0 included basic cloud architecture, key user profiles and their individual workspaces, specialized medical record profile, gene and polymorphism dataset, statistical and image analysis tools. While we will continue to support, improve, and leverage the latest enterprise technologies to further improve performance and scalability, we also have conceptual improvements for a 2.0 release.

Specifically, we will (1) adopt the concept for every Cloud MRM user; (2) provide a convenient user self-registration mechanism; (3) allow customizable, rule based workflows; (4) create individual communication agents to accommodate the array of available institutional data sources; (5) create quick links to a comprehensive disease and gene database; (6) create customizable alerts and triggers; (7) add a 3D data visualization tool to provide visual projection of the determined “trouble spots” onto a human body, (8) add agents in the system to send requests to other users for access privileges on their records; and (9) open API interface to more closely model today's popular web platforms; (10) add a broadcast module for researchers who can choose to "tune in" to select types of information; (11) develop analytical tools with the ability to test complex combinations of therapies responsible for the substantial improvement in long-term outcome.

With further development of its core and extended services, the Cloud MRM will help find solutions to vitiligo and possibly other orphan diseases. In particular, the Cloud MRM may have a wider application in dermatology, as malignant melanoma, the deadliest form of skin cancer, reportedly consists in inverse relationships with vitiligo [19].

Availability and Requirements

Cloud MRM Project home page: https://saas.cyscom.net/sites/VRF/ Authorization required.
Operating system(s): platform independent. Cloud MRM has been tested on Linux, Windows, and Mac OS X.
Other requirements: requires free MS Silverlight plugin to the standard web browser.
License: free for non-commercial use.

References

1. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc. 2006;13:121–126. doi: 10.1197/jamia.M2025
2. As defined by SEC. 526 OF THE FEDERAL FOOD, DRUG, AND COSMETIC ACT [21 USC 360bb], (2)(B)
3. Mark A DePristo, Eric Banks et all. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43, 491–498 (2011) doi:10.1038/ng.806
4. John W. Davey, Paul A. Hohenlohe, Paul D. Etter, Jason Q. Boone, Julian M. Catchen and Mark L. Blaxter. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature. 2011; Vol. 12, 499-510.
5. Can Alkan, Bradley P. Coe & Evan E. Eichler. Genome structural variation discovery and genotyping. doi:10.1038/nrg2958
6. Baird, N. A. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3, e3376 (2008).
7. http://www.kbioscience.co.uk/pricing/cost%20calculator.html
8. Houwink IJF, Van Luijk SJ, Henneman L, Van der Vleuten C, Dinant GJ, Cornel MC. Genetic educational needs and the role of genetics in primary care: a focus group study with multiple perspectives. BMC Family Practice 2011, 12:5
9. Henneman L, Timmermans DRM, van der Wal G. Public experiences, knowledge and expectations about medical genetics and the use of genetic information. Community Genet. 2004;7:33–43. doi: 10.1159/000080302.
10. Genome Wowser: http://www.research.chop.edu/cbmi/wowser
11. MiGenome: http://finance.yahoo.com/news/Illumina-Slashes-IGS-zacks-2114229383.html?x=0&.v=1
12. NHS Choices: http://www.nhs.uk/Pages/HomePage.aspx; Google Health: http://www.google.com/health; Microsoft HealthVault: www.microsoft.com/en-us/healthvault/; Indivo: www.indivohealth.org
13. New Molecular Entity Approvals for 2010 by the U.S. FDA http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/DrugandBiologicApprovalReports/ucm242674.htm
14. Patients Like Me: www.patientslikeme.com; Vitinomics: www.vitinomics.net
15. Christian Vecchiola, Suraj Pandey, Rajkumar Buyya. High-Performance Cloud Computing: A View of Scientific Applications. 2009. ISBN: 978-0-7695-3908-9

16. Arnon Rosenthal, Peter Mork, Maya Hao Li, Jean Stanford, David Koester, Patti Reynolds. Cloud computing: A new business paradigm for biomedical information sharing. Journal of Biomedical Informatics. Volume 43, Issue 2, April 2010, Pages 342-353. doi:10.1016/j.jbi.2009.08.014
17. Amazon 3S: http://aws.amazon.com/s3/; Microsoft Azure Clouds http://www.microsoft.com/windowsazure/
18. Kohane IS, Altman RB. Health-information altruists-a potentially critical resource. N Engl J Med. 2005;353:2074–2077. doi: 10.1056/NEJMsb051220.
19. Spritz RA. The genetics of generalized vitiligo: autoimmune pathways and an inverse relationship with malignant melanoma. Genome Med. 2010. 19;2(10):78.
20. International HapMap Project: http://hapmap.ncbi.nlm.nih.gov/
21. DES Modes of Operation http://www.itl.nist.gov/fipspubs/fip81.htm
22. Summary of the NHIN Prototype Architecture Contracts: A Report for the Office of the National Coordinator for Health IT http://www.hhs.gov/healthit/healthnetwork/resources/summary_report_on_nhin_Prototype_architectures.pdf
23. Garfinkel SL, Shelat A. IEEE Security & Privacy: Data Forensics – Rememberance of Data Passed: A Study of Disk Sanitization Practices. IEEE Distributed Systems Online. 2003;4
24. Whitton ME, Pinart M, Batchelor J, Lushey C, Leonardi-Bee J, González U. Interventions for vitiligo. Cochrane Database of Systematic Reviews 2010, Issue 1. Art. No.: CD003263. doi: 10.1002/14651858.CD003263.pub4.