vcf2fhir: Bridging the Gap Between Genomics and Healthcare

> Dr Robert Dolin Dr Gil Alterovitz Dr Srikar Chamala



vcf2fhir: BRIDGING THE GAP BETWEEN GENOMICS AND HEALTHCARE

On molecular scales, the responses of our bodies to particular medical treatments are deeply engrained in our unique genetic codes. Yet so far, the advanced computer science technologies used to study patient responses and molecular-scale mechanisms have remained entirely independent from each other. Now, **Dr Robert Dolin** of Elimu Informatics, **Dr Srikar Chamala** at the University of Florida, and **Dr Gil Alterovitz** at Brigham and Women's Hospital, address this issue through vcf2fhir: a resource capable of converting between the file formats used by both fields. Through future improvements, his team's approach could soon transform the ways in which crucial clinical decisions are made.



Informatics

Today's technologies allow us to apply the data we gather from many different sources across numerous sectors of society. Such actions encompass an extensive range of fields, including artificial intelligence, robotics, and statistical analysis – yet as a whole, they can be described as part of a broader field, named 'informatics'. As one of the most important branches of modern computer science, this area of technological innovation ensures the continuing operation of many of the data-driven systems we have come to rely on.

Two specific branches of informatics are now key elements of genetic research, and digitised healthcare systems, respectively. On one hand, the field of 'bioinformatics' develops software that can interpret the deeply complex biological datasets describing living organisms. On the other, 'clinical informatics' deals with the application of data to assessing medical problems – allowing clinicians to make more informed decisions about diagnoses and treatments. Yet despite the fact that both fields appear to be highly relevant to one another, and could stand to benefit from each other's operations, there have been few efforts to unify them so far.

Genomics and Molecular Medicine

In modern healthcare systems, clinical informatics can be easily utilised using digital collections of medical information, named Electronic Health Records (EHRs). These records can contain systemised data relevant to both individual patients and wider populations, and can be freely shared across all relevant groups within healthcare systems – ensuring that medical procedures can be carried out as effectively as possible.

In a parallel field of research, the data gathered through bioinformatics can describe precisely how certain medical procedures will affect patients on a molecular level. As a result, this data would be immensely valuable to include in EHRs.



Two particular branches of science would be particularly useful. Firstly, the field of 'genomics' can map out a complete set of a patient's DNA, describing how all of their interrelated genes will collectively respond to certain treatments. Secondly, 'molecular medicine' explores how genomes will respond to the molecular structures and mechanisms contained in medicines – potentially allowing for highly targeted treatments.



'Precision medicine aims to bring together all possible data sources to guide the care of an individual,' describes Dr Robert Dolin of Elimu Informatics in California. 'We are seeing tremendous growth in genomics and molecular medicine – these fields have such voluminous data that they pose challenges for today's EHRs.'

Dr Dolin and his colleagues, Dr Srikar Chamala at the University of Florida, and Dr Gil Alterovitz at Brigham and Women's Hospital in Massachusetts, aim to finally bridge the gap between bioinformatics and clinical informatics. To do this, they look to the latest advances in computer science.

Converting Between File Formats

Many of us will be familiar with the experience of converting files between different formats – whether saving a Word file as a PDF, or converting between JPG and PNG images. To allow us to do this, software developers have designed specialised programs that can essentially express the code underlying a file in a different language. Such translations enable programs that have been built to operate using one file format to handle those initially written in other formats – making them crucial to ensuring that many digital systems can run smoothly.

Through their work, Drs Dolin, Chamala and Alterovitz aimed to build software for carrying out similar conversions between file formats, albeit on colossal scales. In the language of bioinformatics, the data describing an organism's unique genome is stored on a text file in a Variant Call Format (VCF) – which can be readily written, stored, and read out by geneticists.

In clinical informatics, a format named Fast Healthcare Interoperability Resources (FHIR) is used as a common standard for the files used in EHRs. This format stores data in ways that all healthcare providers can work with universally. Ultimately, the differences between these two formats are at the root of the difficulty in unifying bioinformatics and clinical informatics. However, as Dr Dolin and his colleagues have shown, the challenge is not insurmountable.

Bridging the Data Gap

In their latest research, Drs Dolin, Chamala and Alterovitz introduce an advanced piece of software named 'vcf2fhir' - which provides a robust way for researchers, healthcare providers, and any other relevant groups to reliably convert files in the VCF format into FHIR, and vice versa. 'Our vcf2fhir converter is a bridge between molecular data and the EHR,' Dr Dolin explains. 'The converter can extract relevant slices of molecular data, package it up in a language the EHR understands, and deliver it to the EHR, where it can be used by precision medicine algorithms to improve patient care.'

Such advanced capabilities will be immensely useful to clinicians, who likely don't have an in-depth knowledge of the precise molecularscale mechanisms that take place when certain treatments are applied. Using vcf2fhir, they will be able to draw from the findings of cutting-edge research in genomics and molecular medicine – finally bridging the gap between the bioinformatics and clinical informatics communities.

399666367913617976176



In turn, two fields that have developed entirely independently of each other so far will be able to work more closely together – opening up advanced new capabilities in both healthcare and research. Having developed vcf2fhir, Drs Dolin, Chamala and Alterovitz next aimed to test their converter's ability to provide useful guidance in real-world healthcare scenarios.

Practical Decision Making

Perhaps one of the most immediately applicable areas of vcf2fhir is in 'pharmacogenomics' – a field that studies the role of a patient's genome in their response to certain drugs. This is particularly important to consider when predicting allergic reactions in patients. Depending on their unique genome, their reaction of a patient to a particular chemical can cause damaging side-effects, which may appear unexpectedly if the patient hasn't received that treatment before.

Some recent studies have discovered that as many as 7% of all FDA-approved treatments in the US, and some 18% of all written prescriptions, are affected by patient pharmacogenomics. This is now driving a crucial need for clinicians to easily identify how drugs will affect patients on a molecular level, before deciding to prescribe them. In their study, Drs Dolin, Chamala and Alterovitz tested a functional prototype of vcf2fhir for its ability to develop a reliable clinical decision support service – where genomic data could be used to make practical decisions through patient EHRs.

Assessing Drug-gene Interactions

In many medical procedures, it is common for clinicians to require certain information about a patient's genetic code, which may have never been recorded before. In these cases, a conversion from the patient's FHIR files into a VCF format could allow them to determine the exact genetic information they need to obtain to carry out the procedure effectively. By interfacing the vcf2fhir software with patient EHRs, such important clinical decisions could be made quickly and easily, allowing for treatments which are specifically tailored to the pharmacogenomic needs of patients. 'We showed that we could provide drug-gene interaction checking to clinicians right when they are ordering a medication, thereby avoiding potentially serious drug reactions,' says Dr Dolin. Leading on from their initial research, Drs Dolin, Chamala and Alterovitz have now applied vcf2fhir in two further case studies.

Firstly, the SMART Cancer Navigator is a web application that can link patient EHR data with information describing the genetics of various types of cancer. This enables far more coordination between the many medical groups involved in the treatment of cancer patients – leading to more beneficial clinical decisions in turn.

Secondly, the Precision Genomics Integration Platform enables clinicians to intersect a patient's clinical and genomic data with their own knowledge – allowing for the efficient delivery of new, relevant genomic findings to the patient's EHR. From the success of these initial case studies, Drs Dolin, Chamala and Alterovitz now hope that the use of vcf2fhir could be greatly expanded in the near future.

A Bright Future

As an open-source software facility, vcf2fhir could soon become widely accessible, and has already attracted early interest from a wide variety of healthcare institutions across the US. 'Project collaborators come from the University of Florida, Intermountain Healthcare, Cincinnati Children's Hospital, Boston Children's Hospital, Brigham and Women's Hospital, the Harvard/MIT Division of Health Sciences and Technology, and Harvard Medical School,' Dr Dolin concludes.

For now, additional testing will be crucial before vcf2fhir can be realistically applied in real-world healthcare and genomics settings. Yet through this further research, Drs Dolin, Chamala and Alterovitz are hopeful that future efforts to bridge the long-present gap between both fields could lead to profound improvements in the ways that they both operate.







Meet the researchers

Dr Robert Dolin Elimu Informatics Richmond, CA USA **Dr Gil Alterovitz** Brigham and Women's Hospital Boston, MA USA

Dr Robert Dolin completed his MD at the University of California, Irvine in 1986. Ever since, his career has been driven by a goal to improve healthcare through the use of computer technology. In 1989, he became the Chief Resident in Internal Medicine at UCLA, where he designed and implemented the first adaptive Electronic Health Record for patient data. Afterwards, he was instrumental in the development of the Clinical Document Architecture standard; and also developed an interest in genetics along the way. Dr Dolin joined Elimu Informatics in 2009, where he now works as a Senior Informaticist. He now strives to continue his efforts to developing robust clinical standards, and using them to integrate genetic information into Electronic Health Records.

Dr Gil Alterovitz completed his PhD in Electrical and Biomedical Engineering at MIT and Harvard University in 2006. He has since taken research positions at institutions including Harvard Medical School, and the Children's Hospital Boston - before becoming a Lead Investigator at the Brigham and Women's Hospital in 2020. Dr Alterovitz's main research interests focus on the development of novel, interdisciplinary approaches for machine learning in computational biomedicine for infectious diseases. This has led him to develop new ways for studying aspects including drug resistance in tuberculosis, and single-nucleotide polymorphism in DNA.

CONTACT

E: ga@alum.mit.edu W: https://scholar.harvard.edu/gil/ home

WWW.SCIENTIA.GLOBAL

Dr Srikar Chamala University of Florida Gainesville, FL USA

Dr Chamala completed his PhD focusing on Genomics and Bioinformatics at the University of Florida in 2014. In 2017, he became Director of Biomedical Informatics at the University of Florida College of Medicine, where his work focuses on various aspects of genomics, pathology, and clinical informatics. One of Dr Chamala's main research interests is developing informatics strategies for the effective implementation of precision cancer medicine – which has involved efforts including bioinformatics data analysis, and solutions for integrating genomic data into health information systems.

CONTACT

E: schamala@ufl.edu W: https://chamalalab.org/

CONTACT

E: bdolin@elimu.io W: https://www.elimu.io/

FURTHER READING

RH Dolin, SR Gothi, A Boxwala, BSE Heale, A Husami, J Jones, H Khangar, S Londhe, F Naeymi-Rad, S Rao, B Rapchak, J Shalaby, V Suraj, N Xie, S Chamala, G Alterovitz, vcf2fhir: a utility to convert VCF files into HL7 FHIR format for genomics-EHR integration, BMC Bioinformatics, 2021, 22, DOI: 10.1186/s12859-021-04039-1.

RH Dolin, A Boxwala, J Shalaby, A pharmacogenomics clinical decision support service based on FHIR and CDS Hooks, Methods of Information in Medicine, 2018, 57, e115.

7369298



UF FLORIDA

