ABSTRACT

Electronic medical record (EMR)-based precision oncology is a vision that is so far limited to a few pilot and basket studies, with the goal being the design of a proper treatment for cancer patients in real time, based on the panomics knowledge of the patient, and that of similar types of patients. It aims to deliver better treatment outcomes through the design of rational drug combinations, a lower number of futile therapies, reduced patient discomfort, and a healthy human society with a reduced risk of cancer.

The concept of precision oncology began with a few cancer awareness programmes and preventative screenings almost a decade ago. However, the technique took an astronomical leap with the start of the Precision Medicine Initiative Cohort Program and Cancer Moonshot programme very recently. Both projects have invested heavily towards several goals, including the merging of cancer registries and EMR to find the best treatment options for a cancer patient, an idea which, if extended globally, will generate unprecedented possibilities for precision oncology. EMR serve as a broad platform merging a variety of patient information and expert advice to facilitate co-ordinated cancer care. In this article, a summary of the recent EMR-based precision oncology practices for prevention, diagnosis, prognosis, prediction, and their associated concerns and limitations is presented. Though the path of precision oncology is unchartered, the usefulness of real-time information derived from EMR or electronic health records will lead to better precision decision-based oncotherapies.

Keywords: Electronic health records (EHR), electronic medical record (EMR), individualised medicine, panomics, precision cancer medicine, precision decision, stratified medicine.

INTRODUCTION

Cancer is a universal phenomenon, yet each cancer type differs from the other variants. In addition, individual human beings differ from one another by a genome difference of approximately 1%. This tiny difference in the genome sequence, in turn, causes a huge discordance in cancer treatment and drug resistance. The dimensionality of cancer networks is much higher compared to other disease systems; cancer is a genomic disease, harbouring a number of mutated oncogenes and tumour suppressor genes that specify the molecular pathways for tumour growth, sustenance, and progression. The network function depends on the individual patient, external factors, and the ability of the network to overcome treatment mechanisms. The treatment of cancer requires a combined effort among a varied team of experts, including counsellors, medical doctors, caregivers, pathologists, radiologists, surgeons, and others. Irrespective of best efforts, a complex disease like cancer is often unmanageable due to the involvement of genetic, lifestyle, and environmental factors. However, with the use of precision medicine, the situation is improving, albeit slowly. Even though a definitive cure for cancer has not been found, the life expectancy of certain cancer patients has certainly increased.

Predictive, preventative, personalised, and participatory (known as P4) treatment of each individual patient, often coupled with population perspective, is one of the definitions of precision
In simpler words, precision oncology can be defined as diagnosis, prognosis, prevention, treatment, and follow-up that are tailored specifically to an individual patient based on their genetic profile and supported by the genetic profiles of similar types of patients. It is an emerging approach for disease treatment and prevention that takes into account the genetic variability, environment, and lifestyle-related habits of each person.

**DISCUSSION**

Precision oncology is applicable to various aspects of human health, for example, identification of presymptomatic individuals, detection of disease, patient-specific therapy, the study of disease evolution, the design of effective surveillance, and management techniques, among others; altogether, it focusses on providing a uniform platform for human health management.

Traditional evidence-based medical practice along with precision medicine will pave the way for the achievement of such ambitious goals.

The idea of precision oncology gained momentum after a few ‘fit for all’ solutions, i.e., molecularly driven therapies that have been devised for certain types of cancers. At present, imatinib has dramatically improved the survival of chronic myeloid leukaemia patients and HER2-targeted therapies have improved survival rates of women with metastatic HER2-positive breast cancer.

More detailed applications of precision oncology have been reported by Hyman et al. Precision oncology aims to reduce morbidity and mortality, along with improving clinical outcomes of cancer patients, as each patient and cancer type is unique, and so must be the precision decision for each treatment.

Catalogue maintenance and curation of cancer sequence variants are crucial for precision oncology and have immense clinical relevance. Profiling the somatic mutations of genes may predict tumour evolution, prognostics, and treatment. In this context, Tsang et al. reviewed various functional annotation resources and tools to interpret variants in precision genomic oncology. OncoPaD, a web-based tool, has been developed for the rational design of cancer next-generation sequencing (NGS) panels based on mutational data. Based on 7,298 cancer cases from 29 cancer types, it estimates the cost-effectiveness of the designed panel on a cohort of tumours and provides reports on the importance of individual mutations for oncogenesis and therapy. The tool helps in addressing many translational and basic research questions, including those regarding precision oncology. The advent of this NGS technique has a crucial role in precision cancer therapy, a subject that cannot be covered in the present article’s scope, but is well described elsewhere.

However, the idea of large-scale precision oncology is burdened by many hurdles, for example, acquired resistance of tumours, cost-benefit analysis, clinical evaluation, clinical trial design, drug resistance, genetic and genomic analyses of tumours of patients, informatics methods to evaluate genetic variants, liquid biopsy data-related issues, target prioritisation, molecularly targeted drug effectiveness, and tumour heterogeneity, among others. Improvement is needed in the areas of genome sequencing of cancer patients, identification of targets, better access to clinical studies, and design of therapies. In addition, involvement of appropriate parties and stakeholders at every step is crucial for progress in precision oncology, starting from the patient, clinician, physician, caregiver, and third-party insurance providers. These issues should be properly addressed before precision oncology becomes a universal phenomenon for cancer treatment.

In this context, electronic medical records (EMR), the digital version of patient health records, play a crucial role in providing a means by which a patient’s previous health history or health history of patients with similar cancer types can be studied and used for real-time treatment. EMR serve as a broad platform that merges a variety of patient information and expert advice to facilitate co-ordinated cancer care. They aid in personalised cancer risk management and treatment by taking into account the behavioural and social contexts.

An EMR-adapted healthcare system efficiently streamlines day-to-day hospital procedures, along with internal quality control checks, and facilitates easy access to patient data from anywhere around
the globe following proper authentication from a user. EMR are changing the current practice of medicine, as expected almost a decade ago.\textsuperscript{11,25,26}

Interestingly, the basket studies into the use of patient EMR for precision oncology\textsuperscript{12-15,21-23,26-35} have provided meaningful observations for better clinical outcomes. New knowledge developed from the large range of EMR data, with help from predictive models, has generated confidence for precision oncology-based decisions. Precision oncology is a nascent area of therapy but, gradually, with help from information technology and more EMR data, it will take shape from the variety of heterogeneous cancer-related patient data. Precision medicine alongside patient EMR will aid prevention, diagnosis, prognosis, and prediction of cancer types, often one form of assistance combined with others, for the improvement of cancer treatment. Here, we discuss the research developments, or rather the lack of them, in these specified areas.

Preventive Approach

There is currently little scope for adding any cancer prevention measures to a patient’s EMR, as hardly any such measures exist. Mostly, the preventive measures are limited to small cohort screenings\textsuperscript{22} but, as a Pre-Cancer Genome Atlas (PCGA)\textsuperscript{36} is being developed, it will be important to add information on preventative measures into EMR to help in individualised cancer care. For example, programmes like the Million Veterans Program\textsuperscript{37} and the Precision Medicine Initiative Cohort Program\textsuperscript{38} are adding EMR data into their repositories as part of their quest for cancer prevention. Similarly, there are many existing cancer data and knowledge-bases (Table 1) that will be necessary to include for the development of preventative, diagnostic, and prognostic approaches for cancer.

Cancer Diagnosis

To date, the conventional method for cancer diagnosis is mostly based on histopathological reports and radiographic findings. However, the essential role of genomics in the form of biomarkers, molecular signatures, and the molecular profiling reports to detect cancer types cannot be negated. This information is now being stored in the EMR of cancer patients and, in the future, can be used for fast cancer categorisation. Precision cancer diagnostics extend beyond the conventional methods of gene expression profiling and EMR-based decisions. In this respect, we herein discuss some innovations that generate standardised onco-data for inclusion in EMR or electronic health records (EHR).

Pyradiomics\textsuperscript{39} is a Python programming language-based open-source radiomic quantification platform to extract radiomic features from computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) images. It establishes a reference standard for radiomic analyses with immense significance that will reflect in cancer diagnosis.\textsuperscript{40} The Health Level Seven\textsuperscript{®} (HL7) Clinical Genomics group\textsuperscript{41} is working towards the standardisation of genome data and harmonisation with EMR. LesionTracker\textsuperscript{42} is a web-based open-source image assessment and tracking platform for oncology clinical trial workflows. Tumor Map\textsuperscript{43} is an integrated portal for exploration and interrogation of cancer genomics data by using Google’s map technology. It can also facilitate identification of cancer subtypes in silico based on common molecular activities in a set of tumour samples.\textsuperscript{44} The inclusion of such data with EMR or EHR will be beneficial for further analyses at a later stage for diagnostic purposes.

Prognostic Approach

Nowadays, a few rapid learning health systems\textsuperscript{45-48} are in practice for sourcing prognostic and associative factors from text-based cancer EMR. Text-based Exploratory Pattern Analyser for Prognosticator and Associator discovery (TEPAPA)\textsuperscript{45} is a feature learning pipeline for identifying disease-associated factors from unstructured EMR narratives. It has already been applied to a cohort of 82 head and neck squamous cell carcinoma patients’ EMR for phenotyping. CancerLinQ\textsuperscript{174,46} is a cancer decision support system, uses more than a million EMR for any kind of precision decision; it stores data from any format including EMR, provides real-time clinical decision support, measures clinical performance, and generates hypotheses from clinical data to provide lessons learned about technical and logistical challenges. Cancer Commons\textsuperscript{49} is a non-profit network of patients, physicians, and scientists who help in identifying the best options for treating an individual’s cancer. It facilitates rapid learning for precision oncology by gathering panomic and clinical data about individual patients from the Donate Your Data (DYD) registry.

Furthermore, the Melanoma Rapid Learning Utility (MRLU)\textsuperscript{47} is an analytical engine and user interface that provides clinical decision support for
melanoma treatment by identifying and analysing cohorts from a database of 237 metastatic melanoma patients. Precision Oncology 3.0 advocates the analysis of panomic data to hypothesise the patient-specific cancer molecular pathways and their personalised treatments by combining the targeted therapies, and Watson, an IBM developed machine learning system, recommends trials and individualised therapies based on cancer patient data obtained from various sources, including EMR. The Electronic Medical Records and Genomics (eMERGE) Network is also helping to develop the EMR-guided cancer genomic studies and associated decisions.

In addition, cancer-associated research often reflects important information for the oncology community based on EMR data, a few of which are discussed. Gregg et al. performed a study of 2,352 prostate cancer patients' EMR to automate the determination of risk strata according to the D'Amico Risk Classification, in which prostate cancer patients can be labelled as low, intermediate, or high-risk. The natural language processing (NLP) algorithms developed by Gregg et al. had 91% raw agreement with manual risk categorisation, with a recall of 78% (N=1,833). With proper standardisation, such automating procedures for collecting risk characteristics will reduce the time and effort taken, while integrating the information into the respective EHR or disease registries.

Symmans et al. have assessed the long-term prognostic risk of breast cancer patients by analysing 1,158 patients' EMR belonging to five cohorts after neoadjuvant chemotherapy alone or in conjunction with HER-2-targeted treatment. Patient-specific age, classification of primary cancer (whether multifocal or multicentric), clinical and radiologic characteristics of primary cancer, clinical stage of cancer before treatment, diabetic status (present or absent), follow-up for relapse or death, height, and weight data have been considered from the respective EMR for the prognostic study. The study indicates long-term survival after neoadjuvant chemotherapy in all the considered phenotypic subsets of breast cancer, albeit with a need for further external validation.

Heo et al. performed a study of 44 advanced biliary tract cancer patients' EMR to investigate the significance of the clinical impact of overexpression of the proto-oncogene c-MET. Various patient-specific details have been taken from EMR and expression of the c-MET protein has been generated from tumour samples of metastatic patients for this study. In approximately one-third of the patient population, overexpression of c-MET was found, but it cannot be significantly correlated with the other associated variables, generating a need for further studies.

A research study of 99 smokers' EHR performed by Begnaud et al. reflected a randomised electronic promotion of lung cancer screening and Afghahi et al. found the differential use of chemotherapy associated with sociodemographic characteristics by a study of EMR, cancer registry, and genomic data belonging to 293 breast cancer patients. In addition, Rioth et al. created an automated, real-time database of annotated tumour variants of cancer patients to gather data generated during cancer care for secondary use during the precision decision. Schwaederle et al. generated a study using NGS of 439 diverse cancer patients where they integrated the NGS data with age, cancer histology, sex, stage of metastasis at diagnosis (if any), metastatic disease at the time of the biopsy (if any), and race extracted from electronic medical charts, among others, to find actionable aberrations in patients. Lastly, Servant et al. developed a knowledge data interface that facilitates data integration and tracks the processing of individual samples for precision decision in real-time.

As EMR and/or EHR are being considered as a global data source for cancer prognosis, their completeness in accumulating all kinds of patient-related data is necessary. At present, there are many cancer resources (Table 1) and their inclusion or association with the EMR will ensure a more detailed cancer prognosis.

Predicting Cancer Risk

EMR, the data flat files for cancer patients, can be subjected to standard statistical analysis techniques. Symmans et al. used the Kaplan–Meier estimator with 95% confidence intervals estimated using the Greenwood formula with log-log transformation for determination of survival probability. In addition, Heo et al. used the t-test, Fisher's exact test, or one-way analysis of variance, as appropriate, to analyse correlations between expression of the proto-oncogene c-MET and the associated clinicopathologic variables. They used the Kaplan–Meier estimator for the analysis of all time-to-event variables.
### Table 1: Examples of existing cancer data and knowledge-bases.

<table>
<thead>
<tr>
<th>Name</th>
<th>URL</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer Genome Interpreter</strong></td>
<td><a href="https://www.cancergenomeinterpreter.org/home">https://www.cancergenomeinterpreter.org/home</a></td>
<td>Relies on existing knowledge for the identification of alterations in a cancer and detection of therapeutically actionable alterations.</td>
</tr>
<tr>
<td><strong>Cancer Moonshot</strong></td>
<td><a href="https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative">www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative</a></td>
<td>The government of the USA has plans for authorising $1.8 billion in funding for the Cancer Moonshot programme over 7 years, starting from 2017. The project is aimed at vaccine development for cancer types.</td>
</tr>
<tr>
<td><strong>Cancer Driver Log (CanDL)</strong></td>
<td><a href="https://candl.osu.edu/">https://candl.osu.edu/</a></td>
<td>An expert-curated database that lists all possible nucleotide positions for each amino acid change along with their associated literature reference and the chromosome locations.</td>
</tr>
<tr>
<td>cBIOPortal</td>
<td><a href="http://www.cbioportal.org/">http://www.cbioportal.org/</a></td>
<td>Large-scale cancer genomics data sets can be visualised, analysed, and downloaded from this source.</td>
</tr>
<tr>
<td><strong>Clinical Interpretations of Variations in Cancer (CIViC)</strong></td>
<td><a href="https://civic.genome.wustl.edu/">https://civic.genome.wustl.edu/</a></td>
<td>A cost-free source for variant information and cancer-related mutations.</td>
</tr>
<tr>
<td><strong>Genomics Evidence Neoplasia Information Exchange (GENIE)</strong></td>
<td><a href="http://www.aacr.org/Research/Research/Pages/aacr-project-genie.aspx#WZWgmIejHIV">http://www.aacr.org/Research/Research/Pages/aacr-project-genie.aspx#WZWgmIejHIV</a></td>
<td>This project aggregates and links clinical-grade cancer genomic data with clinical outcomes from tens of thousands of cancer patients treated at multiple international institutions.</td>
</tr>
<tr>
<td><strong>Georgetown Database of Cancer (G-DOC)</strong></td>
<td><a href="https://gdoc.georgetown.edu/gdoc/">https://gdoc.georgetown.edu/gdoc/</a></td>
<td>A precision medicine platform containing molecular and clinical data from thousands of patients and cell lines, along with tools for analysis and data visualisation.</td>
</tr>
<tr>
<td><strong>International Cancer Genome Consortium (ICGC)</strong></td>
<td><a href="http://icgc.org/">http://icgc.org/</a></td>
<td>With 89 committed projects, the goal of the ICGC is to obtain a comprehensive description of genomic, transcriptomic, and epigenomic changes for different tumour types, which has great importance across the globe.</td>
</tr>
<tr>
<td><strong>My Cancer Genome</strong></td>
<td><a href="https://www.mycancergenome.org/">https://www.mycancergenome.org/</a></td>
<td>A tool that matches tumour mutations to therapies, making information accessible and convenient for clinicians.</td>
</tr>
<tr>
<td><strong>Personalised cancer therapy</strong></td>
<td><a href="https://pct.mdanderson.org/">https://pct.mdanderson.org/</a></td>
<td>An integrated research and clinical trials strategy for implementing personalised cancer therapy and improving patient outcomes.</td>
</tr>
<tr>
<td><strong>The Cancer Genome Atlas (TCGA)</strong></td>
<td><a href="https://cancergenome.nih.gov/">https://cancergenome.nih.gov/</a></td>
<td>Maps of the key genomic changes in 33 types of cancer have been generated and stored in this database.</td>
</tr>
<tr>
<td><strong>Cistrome Cancer</strong></td>
<td><a href="http://cistrome.org/CistromeCancer/CancerTarget/">http://cistrome.org/CistromeCancer/CancerTarget/</a></td>
<td>Predicted transcription factor targets and enhancer profiles in cancers have been collated from TCGA expression and public ChIP-seq profiles and stored in this resource.</td>
</tr>
</tbody>
</table>
Statistical analysis by Begnaud et al. included the Wilcoxon rank sum test for quantitative characteristics and Fisher’s exact test for categorical items, and the multivariable logistic regression technique was used by Afghahi et al.

EMR data can also be subjected to advanced machine learning techniques for better retrieval of insightful information. Gregg et al. used natural language processing to achieve >90% accuracy for low, intermediate, and high-risk stratification of localised prostate cancer patients. A naïve Bayes classifier was used by Rion et al. to categorise cancer patient care data into 18 separate clinical groups with 99.37% accuracy, and Schweaderle et al. used linear or binary logistic regression analyses and Spearman’s rho coefficients. However, application of these statistical and machine learning techniques in one or more cancer resources (Table 1) is yet to be attempted.

Limitations

Precision oncology has many limitations at the preventive, diagnostic, prognostic, and predictive stages. Molecular profiling of tumours is not yet used for all types of cancer and many of the molecular alterations observed in patients do not respond to drugs; therefore, more research is required regarding the complete characterisation of viable target biomolecules. Diversity also exists among the types of clinical molecular tests and opinions vary among clinical oncologists regarding the interpretation of genomic test results. These inconsistencies will be reflected in the EHR and will diminish confidence in precision oncology-associated therapies.

Most often, DNA-based alterations, such as mutations and gene duplications, among others, are found to be responsible for the development of a typical subtype of cancer in a patient. A large proportion of EMR lack DNA alteration data of cancer patients, since third-party payers rarely reimburse for mutational analysis, and so forms a major bottleneck for progress in precision oncology. Even when the mutations are available, there is no universal guideline for structuring EMR or EHR. Moreover, searching for a common factor across related EHR may be a daunting and large task due to a lack of interconnections, mutual compatibility, and interoperability. If similar types of EHR cannot be analysed in real time, finding a common clue for a disease will prove difficult. To date, no uniform approach to address all the above-mentioned issues is in place. Altogether, the lack of completeness, consistency, homogeneity, and accuracy of cancer EMR or EHR will reflect in precision oncology-based decisions. Hence, a robust universal format and clear regulations should exist for patient EMR to avoid these caveats and to ensure reliability and efficiency.

In developed countries, maintenance of the EHR is a must for the free centralised government healthcare; a report regarding this has been penned by Ben-Assuli. However, in developing and under-developed countries, the maintenance of EHR is performed by third parties, particularly for insurance purposes of the wealthier population who can afford private healthcare facilities. It is often pursued in a very incomplete and sporadic manner. As precision oncotherapies are heavily dependent on population or cohort-based details, EHR data
of developed countries may not help for precision oncology-based treatments in other geographic populations. A lack of proper EHR data will delay and hamper the scope of precision oncology in these countries.

EMR-based precision oncology has provided a large amount of introspection and garnered a lot of focus; however, whether it has generated enough confidence for the decisions to be adopted for clinical and therapeutic purposes is doubtful. A selection of scientists and researchers have already expressed their concerns regarding the performance of precision oncology\(^6^0\)\(^6^1\) and there are doubts regarding clinical usefulness of targeted therapies derived from EMR.\(^6^2\) According to these professionals, current precision oncology practices are too simple an approach for a complex disease like cancer. However, there is also another cohort of professionals who strongly believe in it due to a few clinically useful, although unexpected, advances.\(^6^3\)\(^6^4\) Often, the same researchers are found to sway between feeling hope, optimism, and/or concern regarding precision oncology. Whether the concept of precision oncology will be the next groundbreaking landmark innovation of science, or result in a pseudo-innovation, is not clear.\(^6^5\)\(^6^6\)\(^6^7\)

However, the authors’ opinions favour pragmatic optimism with the advent of the new cloud-based software platforms,\(^6^8\)\(^6^9\) novel ways for phenotypic representation\(^7^0\) and stratification,\(^7^1\) multiomics data analytics platforms,\(^7^2\) and big data infrastructures\(^7^3\) for advancement of precision oncology.

### CONCLUSION

In this paper, the authors have introduced the emerging area of electronic record (EMR/EHR)-based advances for the improvement of cancer treatment. Moreover, this paper has summarised a few of the recent EMR-based precision oncology practices for prevention, diagnosis, prognosis, prediction, and their associated concerns and limitations.

A cancer patient experiences many emotional and physical hurdles, starting from the initial confirmation and acceptance of their diagnosis. The associated stress, stigma, and sickness looms like a black cloud over the patient. Irrespective of any kind of curative or palliative care, a cancer patient’s life expectancy certainly decreases following diagnosis; some patients have a few days of life or a few years, while others live longer, but always with a fear of tumour recurrence. In these scenarios, even a few scientific details derived from EMR and some unexpected advances that aid in oncotherapy become a baton of hope and belief.

The idea of EMR data-derived precision oncology is novel and the path for achieving this idea is unchartered, while the limitations are many. With the increase of lifestyle addictions and disorders coupled with the genetic and environmental factors, cancer is becoming the second leading cause of global death,\(^7^4\) with 75% of deaths occurring in low and middle-income countries. However, the unbeatable human desire for cancer-free, longer, and healthier lifespans also exists and EMR/EHR will lead the way towards this through precision oncology-based therapies.

### REFERENCES

13. Bower H et al. Life expectancy of...


If you would like reprints of any article, contact: +44 (0) 1245 334450.