

Opportunities for Patient Matching Algorithms to Improve Patient Care in Oncology

The Promise of Precision Medicine

The premise of precision medicine is rooted in the hypothesis that diseases are heterogeneous and that each person's disease is unique; therefore, that person needs to be treated as an individual.¹ Conditions for the application of precision medicine in oncology have never seemed better. Since the completion of the Human Genome Project in 2003, we now have several large cancer-specific genomic sequencing and cataloging initiatives, including The Cancer Genome Atlas, the Wellcome Sanger Institute Catalogue of Somatic Mutations in Cancer, and the International Cancer Genome Consortium. In addition, several major efforts have been made to better understand the unique clinical and genomic features that predict response to cancer treatment, such as the National Cancer Institute Exceptional Responders Initiative. Finally, with the advent of electronic health records and with new standards for electronic clinical documentation, is the emergence of clinical data sharing projects, such as the ASCO CancerLinQ (Cancer Learning Intelligence Network for Quality) program, that propose to improve the quality of cancer service delivery through big data science.

With the cost of clinical sequencing continuing to decline² and new mandates for electronic data capture, now is an opportune time for implementing a truly personalized medicine approach to cancer treatment. However, current approaches to personalized medicine in oncology fail to subsume, for the most part, the full complexity of an individual's case when developing predictive and prognostic models and instead focus on a small subset of features (biomarkers). We refer to this approach as the reductionist paradigm: Although it may be appropriate for certain well-defined situations, the approach fails to address large knowledge gaps. Outside a few mutations (eg, epidermal growth factor receptor mutations in non-small-cell lung cancer [NSCLC], *BRAF* V600 mutations in

melanoma), predictive gene expression signatures (eg, OncotypeDX [Genomic Health, Redwood City, CA]^{3,4}), or prognostic nomograms (eg, Kattan nomogram for prostate cancer⁵), clinical application of biomarkers in oncology has been slow to advance and has been stymied by patient heterogeneity; cost; and in part, the curse of dimensionality, which often leads to ambiguous selection of candidate biomarkers from long lists of differentially expressed genes.⁴

Traditional Biomarker Paradigm

A traditional biomarker study begins with a screening process whereby a large panel of potential features (eg, somatic mutations, differentially expressed genes, patient demographics) is compared with a particular phenotype of interest. Through the use of feature selection methods, a subset of biomarkers with the strongest association to the feature of interest is identified and then validated in an independent test set. This systematic reduction in features is what we refer to as the reductionist paradigm and assumes that this limited number of biomarkers can be reliably and consistently used to guide treatment decisions.⁶ Restated, traditional biomarker approaches generally assume that a global, uniform, biologic truth about the drug-patient relationship can be consistently detected.

In the oncology sphere, the most successful applications of the reductionist paradigm have been in the prediction of response to treatment with targeted agents in cancers where the drug target is the major (if not sole) driver of oncogenesis in that cancer. In these cases, detection of an activating mutation in the drug target in an individual patient's cancer is a powerful predictor of response to the targeted therapy. Classic examples include response to imatinib in patients with BCR-ABL translocated chronic myelogenous leukemia,⁷ response to erlotinib in epidermal growth factor receptor mutant NSCLC,⁸ and response to trastuzumab in human epidermal growth factor receptor 2-amplified breast cancer.⁹

Travis Johnson
David Liebner
James L. Chen

The Ohio State University,
Columbus, OH

Corresponding author:
James L. Chen, MD, The
Ohio State University, 250
Lincoln Tower, 1800
Cannon Dr, Columbus, OH
43210; e-mail: james.
chen@osumc.edu.

Unfortunately, most oncology patients' cancers cannot be succinctly described by a single set of biomarkers. Indeed, even for the aforementioned cases, a spectrum of outcomes and responses to any given therapeutic agent is well acknowledged within any one cohort (eg, *BRAF* V600 mutant melanoma).¹⁰ To further tease out the subtleties of patient response, one must incorporate multiple secondary biomarkers. In metastatic melanoma, for example, outcomes for patients with *BRAF*V600 mutations in melanoma treated with targeted therapies have been associated with the number of metastatic sites and serum lactate dehydrogenase levels.¹¹ A tendency occurs, therefore, to divide patients into increasingly smaller subsets on the basis of biomarker status, and this subdivision tends to proceed until further subdivision is statistically untenable. The result is the fragmentation of previously common diseases into a collection of rare subtypes, which are challenging to study, and results in many more drug combinations and drug-gene interactions than can be feasibly assessed.

Clinical Judgment: The Original Similarity-Based Approach

A tenable solution to our modern biomarker problem may be grounded in the past. Before the concept of biomarker-driven precision medicine was clinical judgment, sometimes thought of as clinical gestalt or the idea that medical practitioners can subsume disparate phenomena in a patient into a singular narrative that accurately describes the patient's condition. Clinical judgment is based on a physician's ability to pattern match on sparse unstructured data and is a comprehensive and integrative approach to decision making.¹² Implicit in this belief is the ability for clinicians to handle sparse data to make clinical decisions in the absence of complete information and generate solutions that can be reused.¹³ Indeed, experienced physicians often are sought for their ability to rapidly and accurately diagnose and treat conditions because they had seen similar patients in the past.

Clinical judgment has been compared quantitatively with various standardized tests, such as the HEART score, a metric based on the patient's history, electrocardiogram, age, risk factors, and troponin level. Clinical judgment and the HEART score were comparable in predicting acute coronary syndrome (ACS) risk ($P = .13$).¹⁴ Other studies have compared ACS discharges between clinical judgment and a combined measure of clinical judgment, electrocardiogram findings, and troponin levels.

Clinical judgment had a sensitivity of 0.72 (95% CI, 0.61 to 0.81) and a specificity of 0.69 (95% CI, 0.64 to 0.74). The combined measure had a sensitivity of 1.00 (95% CI, 0.95 to 1.00) and a specificity of 0.47 (95% CI, 0.41 to 0.52).¹⁵ In predicting ACS, the use of standardized tests alone over clinical judgment alone showed no significant improvement. However, the combination of clinical gestalt and standardized tests shows a stark improvement, and in fact, cancer treatment can learn from this model. We advocate that oncologists use standard clinicopathologic metrics in addition to their understanding of a patient's best matches.

Clinical judgment is in direct contrast to reductionist biomarker approaches and has been used effectively since the advent of medicine. One could argue that clinicians intuitively embrace heterogeneity and perform pattern matching at the patient level. In fact, some of the most powerful predictors of patient outcomes are not particular laboratory values but comprehensive and integrative assessments such as the performance status, a measure of a patient's overall well-being through combined quantitative (test values) and qualitative observations (physical examination).¹⁶

Automation of Clinical Judgment: Opportunities in the Electronic Medical Record Era

The incorporation of electronic health records into standard clinical practice provides a prime opportunity for algorithmic approaches to automate clinical judgment, which may provide a living and learning laboratory (Fig 1). As with any new technology, key challenges exist in developing similarity-based algorithms. These challenges fall into three categories: data heterogeneity, algorithm selection, and data sharing (Appendix Table A1, online only).

Challenge 1: Data heterogeneity. The approximation of clinical judgment requires that algorithms subsume disparate data types and comparison methods. A patient's features could potentially be compared not only from patient to patient (eg, Foundation Medicine ICE [Interactive Cancer Explorer] platform) but also from patient to cell line (eg, Connectivity Map project¹⁷), patient to abstract patient archetypes (comparison of patient DNA with a reference genome), and one patient's electronic medical record to that of another.¹⁸ This method is complicated by the data types used in the matching process, which can vary from subjective to objective phenotypic measurements, the vast array of omics assays, incomplete data sets, and data sets that lack standardized ontologies.

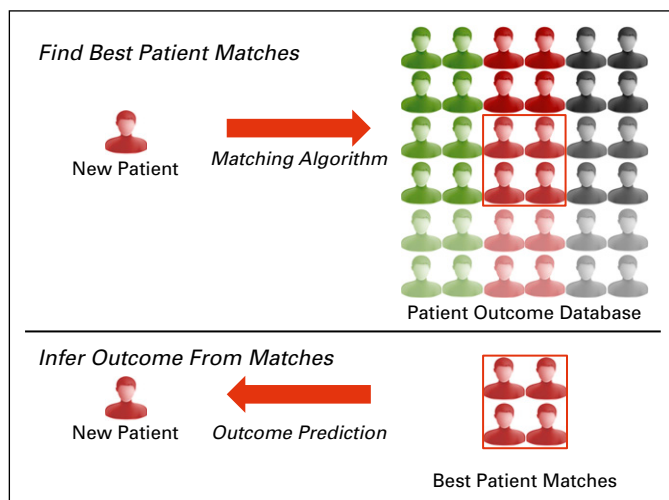


Fig 1 – Matching algorithm paradigm. The information of a new patient is compared with a set of reference patients. The phenotypic outcomes of the best matches of the reference patients is then used to infer the outcomes of the new patient.

Preprocessing steps are needed to transform the data into viable features for use in matching algorithms. The possibility that the best predictive models may require disparate unstandardized data types to be subsumed simultaneously complicates this preprocessing¹⁹⁻²² and requires standardization of data between data sets, including the mapping of unstandardized ontologies and normalization. However, as the technology and standards mature through efforts such as the Global Alliance for Genomics and Health (GA4GH)^{23,24} and the HL7 FHIR (Fast Healthcare Interoperability Resources),^{25,26} such a standards convergence should occur.

Challenge 2: Algorithm selection. This wide array of data leads to equally heterogeneous algorithms and metrics.²⁷⁻³⁰ At the present, matching algorithms can be separated into feature matching and outcome matching algorithms. Feature matching algorithms assume that retained features are critical determinants of outcomes (eg, drug response) and are optimal for situations where the biomarker is directly linked to the outcome. A straightforward approach to feature matching is to assign matches that are based on exact feature overlap; that is, for two patients to be a match, the patients must share all features. The Foundation Medicine Patient-Match tool (Appendix Table A1) is an example of this exact matching approach. More complex feature matching schemes that use Bayesian approaches have been developed.³¹ Other feature matching algorithms include PHIVE (phenotypic interpretation of variants in exomes), which matches human phenotypic profiles to glean the variants found in whole exome sequencing in mouse models,²⁹ and DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans

Using Ensembl Resources), which enables international querying of karyotype and genetic and phenotypic information for matches.³² Aside from matching patients to patients or clinical models, feature matching approaches have been applied to clinical trial recruitment.

The number and scope of clinical trials result in recruitment problems among potential participants. The matching of patients to trials by using clinical trial alerts increases feasibility and sustainability in recruitment of large sample sizes by increasing the number of physicians who generate referrals.³³ Certain clinical trial alerts have shown up to a 90% reduction in workload in trial recommendations.³⁴

In contrast to feature matching, outcome matching allows features to be weighted on the basis of their discriminatory power. Common algorithms, such as weighted k-nearest neighbor, random forest plots, and neural networks, can be used.³⁵⁻³⁷ Allele frequency of single nucleotide variants between patients and NSCLC cell lines can be used to predict chemotherapeutic response to treatment.²⁷ Modeled from online dating, GeneYenta matches phenotypically similar patients with regard to rare diseases³⁶ by weighting predictive features. Other start-up efforts, such as MatchTX,³⁸ are attempting to leverage social networking tools to discover best patient matches by using genomic and phenomic data simply as the intermediary. Thus, although the data sources, data types, and methods are heterogeneous, matching techniques at their core use heuristic approaches to discover and vet the best profiles from large clinical databases and algorithmically use the same dependable methods that have been in use since the advent of medicine but across a far broader integrative knowledge base.

Challenge 3: Data sharing. To acquire the critical mass of data and expertise for patient matching, perhaps the most exciting developments have been initiated by the GA4GH, an international consortium that enables the sharing of genomic and clinical data. The goal of Matchmaker Exchange, a project associated with GA4GH, is to improve the identification of the causal variants of rare diseases through the adoption of GA4GH data standards and data sharing agreements among participant sites.³⁹ The GA4GH Beacon Project allows federated queries to detect the existence of specific genomic variants across a variety of genome resources. ASCO also has built a prototype system, CancerLinQ, to facilitate integration of data from multiple participating community oncology

practice sites to standardize data, facilitate research, and provide personalized cancer care through patient matching.^{40,41} In the United States, academic institutions have been participating in consortia such as ORIEN (Oncology Research Information Exchange Network),^{42,43} GENIE (Genetics of Nephropathy—an International Effort),^{42,44} the International Cancer Genome Consortium,⁴⁵ and OPeN (Oncology Precision Network)⁴⁶ to build their respective frameworks for identifying patient cohorts. Indeed, the ambitious Sync for Science^{47,48} endeavor sponsored by the National Institutes of Health and the Office of the National Coordinator for Health Information Technology will permit patients to directly donate their tissues for precision medicine research to support innovative match-based algorithms for predictive purposes.

The Foundation Medicine ICE platform, Kuveda, and PatientsLikeMe are all commercial efforts in data sharing. Kuveda attempts to automate and distribute personalized profiling and drug choice by creating tools to aid oncologists in genomic profiling of patients. PatientsLikeMe connects patients with one another and collects data from participating patients for use by researchers. These large consortia and businesses all rely on patient matching as part of their core strengths.

In conclusion, the early promise of whole genome sequencing and the reduction of molecular variants into clinically actionable diagnostics have been more difficult than initially anticipated. The direct translation of small subsets of variants has not encompassed the required complexity to interpret heterogeneous diseases, and the algorithms that call these variants are not yet optimized. An alternative to this approach is to focus on the use of complex feature sets that include clinicogenomic features to match patients to patients and then extrapolate information about the disease from the matching patients. Fortunately, such matching tools specifically designed for the genomic era are rapidly becoming commonplace (Appendix Table A2, online only). Taken together, we and others advocate for the embracing of genomic/phenotypic heterogeneity.⁶ Ongoing large-scale data banking efforts, global efforts to standardize omics data, and the development of advanced algorithms to perform sophisticated cohort selection are rapidly enabling patient matches as a viable methodology for prediction. Ultimately, patient matching should result in a sea change in our ability to rapidly integrate precision medicine into clinical practice by computationally enhancing the clinical judgment of our practitioners.

DOI: [10.1200/CCI.16.00042](https://doi.org/10.1200/CCI.16.00042)

Published online on jcocci.ascopubs.org on February 23, 2017.

AUTHOR CONTRIBUTIONS

Conception and design: All authors

Financial support: James L. Chen

Administrative support: James L. Chen

Provision of study material or patients: James L. Chen

Collection and assembly of data: Travis Johnson, James L. Chen

Data analysis and interpretation: James L. Chen

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Travis Johnson

No relationship to disclose

David Liebner

Research Funding: Lilly Oncology, CytRx, Bayer, Novartis, Bristol Myers Squibb, Ignyta, Morphotek, Mirati Therapeutics, Millennium Pharmaceuticals, Merck, Incyte, Sanofi, Vertex Pharmaceuticals, Foundation Medicine

Patents, Royalties, Other Intellectual Property: Methods for Predicting Prognosis (MatchTx): provisional US patent application #14/912,961

James L. Chen

Stock and Other Ownership Interests: MatchTX

Consulting or Advisory Role: Novartis, Eisai

Speakers' Bureau: Novartis

Patents, Royalties, Other Intellectual Property: Methods for Predicting Prognosis (MatchTx): provisional US patent application #14/912,961

REFERENCES

1. Ginsburg GS, McCarthy JJ: Personalized medicine: Revolutionizing drug discovery and patient care. *Trends Biotechnol* 19:491-496, 2001
2. Mardis ER: Anticipating the 1,000 dollar genome. *Genome Biol* 7:112, 2006

3. Paik S, Shak S, Tang G, et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351:2817-2826, 2004
4. Sparano JA, Paik S: Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol* 26:721-728, 2008
5. Kattan MW, Eastham JA, Stapleton AM, et al: A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 90:766-771, 1998
6. Lussier YA, Chen JL: The emergence of genome-based drug repositioning. *Sci Transl Med* 3: 96ps35, 2011
7. Druker BJ, Talpaz M, Resta DJ, et al: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 344:1031-1037, 2001
8. Pérez-Soler R, Chachoua A, Hammond LA, et al: Determinants of tumor response and survival with erlotinib in patients with non-small-cell lung cancer. *J Clin Oncol* 22:3238-3247, 2004
9. Yeon CH, Pegram MD: Anti-erbB-2 antibody trastuzumab in the treatment of HER2-amplified breast cancer. *Invest New Drugs* 23:391-409, 2005
10. Sosman JA, Kim KB, Schuchter L, et al: Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med* 366:707-714, 2012
11. Long GV, Weber JS, Infante JR, et al: Overall survival and durable responses in patients with *BRAF* V600-mutant metastatic melanoma receiving dabrafenib combined with trametinib. *J Clin Oncol* 34:871-878, 2016
12. Mitchell AM, Garvey JL, Chandra A, et al: Prospective multicenter study of quantitative pretest probability assessment to exclude acute coronary syndrome for patients evaluated in emergency department chest pain units. *Ann Emerg Med* 47:447.e1, 2006
13. Kabrhel C, Camargo CA Jr, Goldhaber SZ: Clinical gestalt and the diagnosis of pulmonary embolism: Does experience matter? *Chest* 127:1627-1630, 2005
14. Visser A, Wolthuis A, Breedveld R, et al: HEART score and clinical gestalt have similar diagnostic accuracy for diagnosing ACS in an unselected population of patients with chest pain presenting in the ED. *Emerg Med J* 32:595-600, 2015
15. Body R, Cook G, Burrows G, et al: Can emergency physicians 'rule in' and 'rule out' acute myocardial infarction with clinical judgement? *Emerg Med J* 31:872-876, 2014
16. Donnelly ED, Refaat T, Gentile M, et al: Evaluation of outcomes in patients with carcinoma of the cervix treated with concurrent radiation and cisplatin versus cisplatin/5-FU compared with radiation alone. *Am J Clin Oncol* 38:437-441, 2015
17. Lamb J, Crawford ED, Peck D, et al: The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929-1935, 2006
18. Mate S, Köpcke F, Toddenroth D, et al: Ontology-based data integration between clinical and research systems. *PLoS One* 10:e0116656, 2015 [erratum: *PLoS One* 10:e0122172, 2015]
19. Wang B, Mezlini AM, Demir F, et al: Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 11:333-337, 2014
20. Ritchie MD, Holzinger ER, Li R, et al: Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* 16:85-97, 2015
21. Kristensen VN, Lingjærde OC, Russnes HG, et al: Principles and methods of integrative genomic analyses in cancer. *Nat Rev Cancer* 14:299-313, 2014
22. Shen R, Mo Q, Schultz N, et al: Integrative subtype discovery in glioblastoma using iCluster. *PLoS One* 7:e35236, 2012
23. Lawler M, Siu LL, Rehm HL, et al: All the world 's a stage: Facilitating discovery science and improved cancer care through the global alliance for genomics and health. *Cancer Discov* 5:1133-1136, 2015
24. Global Alliance for Genomics and Health: GENOMICS: A federated ecosystem for sharing genomic, clinical data. *Science* 352:1278-1280, 2016

25. Alterovitz G, Warner J, Zhang P, et al: SMART on FHIR genomics: Facilitating standardized clinico-genomic apps. *J Am Med Inform Assoc* 22:1173-1178, 2015
26. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. *Proc IEEE Int Symp Computer-Based Medical Systems* 326-331, 2013
27. Dudley JT, Chen R, Butte AJ: Matching cancer genomes to established cell lines for personalized oncology. *Pac Symp Biocomput* 243-252, 2011
28. Embi PJ, Jain A, Clark J, et al: Development of an electronic health record-based Clinical Trial Alert system to enhance recruitment at the point of care. *AMIA Annu Symp Proc* 231-235, 2005
29. Robinson PN, Köhler S, Oellrich A, et al: Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* 24:340-348, 2014
30. Wicks P, Vaughan TE, Massagli MP, et al: Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol* 29:411-414, 2011
31. Satagopan JM, Sen A, Zhou Q, et al: Bayes and empirical Bayes methods for reduced rank regression models in matched case-control studies. *Biometrics*, 72:584-595, 2016
32. Firth HV, Richards SM, Bevan AP, et al: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 84:524-533, 2009
33. Embi PJ, Jain A, Clark J, et al: Effect of a clinical trial alert system on physician participation in trial recruitment. *Arch Intern Med* 165:2272-2277, 2005
34. Ni Y, Wright J, Perentesis J, et al: Increasing the efficiency of trial-patient matching: Automated clinical trial eligibility pre-screening for pediatric oncology patients. *BMC Med Inform Decis Mak* 15:28, 2015
35. Chen HL, Huang CC, Yu XG, et al: An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst Appl* 40:263-271, 2013
36. Chen Y, Cao W, Gao X, et al: Predicting postoperative complications of head and neck squamous cell carcinoma in elderly patients using random forest algorithm model. *BMC Med Inform Decis Mak* 15:44, 2015
37. Amato F, López A, Peña-Méndez EM, et al: Artificial neural networks in medical diagnosis. *J Appl Biomed* 11:47-58, 2013
38. MatchTX: MatchTX: Matching patients to cancer treatments. <http://match-tx.com>
39. Philippakis AA, Azzariti DR, Beltran S, et al: The Matchmaker Exchange: A platform for rare disease gene discovery. *Hum Mutat* 36:915-921, 2015
40. Sledge GW, Miller RS, Hauser R: CancerLinQ and the future of cancer care. *Am Soc Clin Oncol Educ Book* 430-434, 2013
41. Schilsky RL, Michels DL, Kearbey AH, et al: Building a rapid learning health care system for oncology: The regulatory framework of CancerLinQ. *J Clin Oncol* 32:2373-2379, 2014
42. Chakradhar S: Group mentality: Determining if targeted treatments really work for cancer. *Nat Med* 22:222-224, 2016
43. The James: 'ORIEN' Precision Cancer Research Collaboration—Coanchored by Ohio State—adds three new partners, 2015. <https://cancer.osu.edu/news-and-media/news/orien-adds-three-leading-oncology-institutions-to-growing-precision-cancer-research-collaboration>
44. American Association for Cancer Research: American Association for Cancer Research launches international genomic and clinical data-sharing project, 2015. <http://www.aacr.org/Newsroom/Pages/News-Release-Detail.aspx?ItemID=781#.WGeg9FMrKM8>
45. Joly Y, Dove ES, Knoppers BM, et al: Data sharing in the post-genomic world: The experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). *PLOS Comput Biol* 8:e1002549, 2012
46. Synapse: Oncology Precision Network (OPeN) announces data sharing commitments at Vice President Biden's Cancer Moonshot Summit, 2016. <http://www.syapse.com/blog/open-press-release-june-2016>

47. Turvey C, Klein D, Fix G, et al: Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. *J Am Med Inform Assoc* 21:657-663, 2014
48. Riley WT, Nilsen WJ, Manolio TA, et al: News from the NIH: Potential contributions of the behavioral and social sciences to the precision medicine initiative. *Transl Behav Med* 5:243-246, 2015

APPENDIX

Table A1 – Brief List of Algorithms, Companies, and Consortia That Use Patient Matching

Source	Description
Algorithms/tools	
GeneYenta	A Web-based tool used to facilitate the sharing of genomic information among rare disease researchers by matching phenotypically similar patients.
Phenotypic Interpretation of Variants in Exomes (PHIVE)	An algorithm that leverages the vast information in mouse models for human disease research by matching human disease to phenotypically similar mouse models.
PatientMatch	A tool developed by Foundation Medicine for Interactive Cancer Explorer 2 (ICE2) that matches genomically similar patients to facilitate knowledge sharing between physicians.
PhenomeCentral	A member of the Matchmaker Exchange that uses both the PHIVE algorithm and the Matchmaker Exchange application programming interface. PhenomeCentral is a restricted-access Web portal for clinicians and researchers to share phenotype and genotype data.
Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER)	A member of Matchmaker Exchange, DECIPHER is an online platform that facilitates genomic and phenomic matching to aid researchers in their identification of possible genomic profiles with similar disease phenotypes.
GeneMatcher	A member of Matchmaker Exchange, GeneMatcher is a Web-based tool for clinicians, researchers, and basic science researchers to discover and collaborate with other researchers with complementary expertise and interest in the same genes.
Cafe Variome	A lightweight software package used to display and/or share biomedical data from participating researchers.
Genesis (formerly GEM.app)	A member of the Matchmaker Exchange, Genesis is a user-friendly software tool for storing, analyzing, and sharing next-generation sequencing data.
GenomeConnect	Developed by ClinGen and a member of Matchmaker Exchange, GenomeConnect is the patient portal provided by ClinGen so that patients can submit detailed data to build patient registries.
Companies	
Foundation Medicine	A Roche company that develops, manufactures, and sells genomic analysis diagnostics. This company is responsible for the ICE2 platform and PatientMatch algorithm (www.foundationmedicine.com).
Kuveda	A cancer health company that uses an analytics platform to recommend the best drug therapy in relation to the patient's molecular profile (www.kuveda.com).
PatientsLikeMe	A company that facilitates information sharing among patients, researchers, and health-related industry partners (www.patientslikeme.com).

(Continued on following page)

Table A1 – Brief List of Algorithms, Companies, and Consortia That Use Patient Matching (continued)

Source	Description
MatchTX	A big data analytics software company that matches patients to cancer treatments by using combined genomic data and clinical outcomes (http://match-tx.com).
Consortia	
Global Alliance for Genomic Health	
Matchmaker Exchange	Matchmaker Exchange is a federated network of databases with the goal to find genetic causes of rare diseases by matching similar phenotypic and genotypic profiles (www.matchmakerexchange.org).
Beacon Project	The Beacon Project is an initiative that brings together international sites to share genetic data by using specific genomic queries by variant and loci (www.ga4gh.org/#/beacon).
The Monarch Initiative	An international consortium that aims to improve deep phenotyping and integration of both human and model organism ontologies (https://monarchinitiative.org).
Undiagnosed Disease Network (UDN)	A member of the Matchmaker Exchange, the UDN facilitates data sharing and uses exchange tools like PhenomeCentral to further its research of undiagnosed disease (www.undiagnosed.hms.harvard.edu).
Cancer Learning Intelligence Network for Quality (CancerLinQ)	CancerLinQ is an ASCO project that collects data from patients within participating sites to improve treatment decisions and research through data collection and patient matching techniques (http://cancerlinq.org).
International Cancer Genome Consortium (ICGC)	ICGC is a collection of projects that aims to produce a comprehensive list of genomic changes in 50 cancer types of interest (http://icgc.org).
The Cancer Genome Atlas (TCGA)	TCGA is a cancer-specific omics and clinical data gathering initiative that allows users to view and download combined omics and clinical data sets (https://cancergenome.nih.gov).
Catalogue of Somatic Mutations in Cancer (COSMIC)	COSMIC is a comprehensive resource that allows the exploration of variant effects on cancer (http://cancer.sanger.ac.uk/cosmic).
Exceptional Responders Initiative (ERI)	ERI aims to understand the underpinnings of exceptional responders to treatment, including chemotherapeutics (www.cancer.gov/news-events/cancer-currents-blog/2015/exceptional-responders-initiative).

Table A2 – Summary of Pros and Cons of Patient Matching Approaches

Pros	Cons
Comprehensive and integrative view of patients in which the unit of measure is the individual patient	Large clinical and genomic databases are relied on for matching.
Inclusion of disparate data types (clinical, molecular, lifestyle)	Standards for data sharing and large consortia are required to govern databases.
Algorithms and databases already exist that can be used in patient matching	Patient matching algorithms form a new development that has not been thoroughly tested.
More cost-effective than the running of large numbers of clinical trials	Data aggregation and sharing costs themselves are costly, and revenue models must be developed to encourage such collaborations.