

Artificial Intelligence toward Personalized Medicine

Muhammad Wildan Gifari¹, Pugud Samodro², Dhadhang Wahyu Kurniawan^{3*}

¹Department of Biomedical Engineering, Institut Teknologi Sumatera, Lampung, Indonesia

²Faculty of Medicine, Universitas Jenderal Soedirman, Purwokerto, Central Java, Indonesia

³Department of Pharmacy, Faculty of Health Sciences, Universitas Jenderal Soedirman, Purwokerto, Central Java, Indonesia

ABSTRACT

In current medical practice when a patient feels symptoms he/she would consult the doctor. The doctor then gives medication in a one-fits-all fashion. However, recent genetics studies had shown that different genetic makeup can result in different effects on medication, so the medication should be customized for every individual. The main idea of “personalized medicine” is to provide the right intervention including medication to the right patient at the right time and dose. With this approach, the medication paradigm would shift from curative to preventive. The rise of personalized medicine had been possible because the information from ever-increasing biomolecular (proteomics, genomics, and other omics) and health-related data are successfully “mined” by Artificial Intelligence (AI) tools. In this paper, we proposed that AI systems toward personalized medicine must have acceptable performance, be readily interpretable by the clinical community, and be validated in a large cohort. We examined a few landmark papers with the keyword “AI for personalized medicine application”; 1) automatic image-based patient classification, 2) automatic gene-based cancer classification, and 3) automatic health-record heart failure with preserved ejection fraction patient phenotyping. All the examples are evaluated by their performance, interpretability, and clinical validity. From the analysis, we concluded that AI for personalized medicine could benefit by five factors: (1) standardization and pooling of genetics and health data, nationally and internationally, (2) the use of multi-modalities data, (3) disease specialist to guide the development of AI model, (4) investigation of AI-finding by clinical community, and (5) follow-up of AI-finding by the large clinical trial.

Keywords: artificial intelligence; personalized medicine

ARTICLE HISTORY

Received : January 2021

Revised : August 2021

Accepted : August 2021

*corresponding author

Email : dhadhang.kurniawan@unsoed.ac.id

INTRODUCTION

In current medical practice, the doctor diagnoses a pathology for patients according to symptoms and clinical test, then prescribes medication in one-fits-all fashion, without considering genetics, metabolomics, or proteomics of the patient. If the patient has side effect with the medication he is taking, then the doctor will adjust the prescription. Although effective for mass manufacturing of drugs, this medication paradigm does not consider the biomolecular “omics” of the patient. Genetic effect of drug response had been demonstrated on racially diverse children with asthma (Mak et al., 2018). Another paper reviewed the genetics effect on diabetes mellitus type 2 prognosis (Ingelsson & McCarthy, 2018). In addition to genomics, metabolomics effects had been shown to affect response of paracetamol in rats (Clayton et al., 2006). Combination of “omics” data could identify genetics difference of patients that respond to Selective Serotonin Reuptake Inhibitors (SSRIs) (Nguyen et al., 2021). These findings that correlate genetic plus “omics” with drug response raised the question whether the medicine could be personalized.

Personalized medicine is a new paradigm in which disease prognosis and medication are monitored based on genomics profile or other health-related markers in personalized way. This is different from evidence-based medicine in which the best medication is based on average treatment effect (Blackstone, 2019). The vision of personalized medicine is to provide right intervention to the right patient, at the right time and dose (Fröhlich et al., 2018). Because disease prognosis and progression can be linked to individual gene and biomarker, medication would shift from reactive, in which drugs and treatments are performed when the patient already showed some symptoms, toward active, in which the doctor administers drugs and treatments at the very onset of the disease, even before the expected disease shows any symptoms. A promising example is the personalized medicine for cancer prevention (Kensler et al., 2016) and bioinformatics for cancer treatment (Singer et al., 2017).

There are some factors that might impact personalized medicine in past decade. One important factor is the gene sequencing advancement. Next generation gene sequencing technology (Shendure & Ji, 2008) and the

price decline of gene sequencing provide rapid and affordable acquisition of genomics information across health system (Abul-husn & Kenny, 2019). Biobanks initiatives to pool genomics data of hundreds of thousands of participants are sprawling in countries such as Denmark (Agerbo et al., 2015), United Kingdom (Collins, 2012), China (Chen et al., 2011), and Japan (Nagai et al., 2017). The emerging field of pharmacogenomics stem on the premise that drug response depends on genetic makeup of the patients.

Other factor that drives personalized treatment is the availability of health record data such as MRI and CT scans, Electronic Medical Records (EMR), health claims, and ambulatory data from mobile sensors that can be utilised for clustering patients group (Hoffman & Williams, 2011; Jensen et al., 2012). In addition, population wide EMR enable discovery of disease progression on subtypes of population in both cross-sectional and longitudinal fashion (Huopaniemi et al., 2014). Furthermore, there is an initiative to collect and systemize both EMR and genomics data in EMR-linked biobanks to facilitate personalized medicine discovery and implementation (Gottesman et al., 2013). With rapidly increasing genetics, omics, and EMR data, scientist needs informatics tool to manage and extract useful information.

Indeed, the rise of personalized medicine would not be possible without Artificial Intelligence (AI) tool to extract information from the pool of data. Specifically, AI tool called Machine Learning (ML) is utilised to find correlation and pattern in the data using statistical inference. Machine Learning algorithm can be categorized into supervised and unsupervised learning. In the supervised learning, data is divided into training set which has the correct output labeled by human “supervisor” to create a predictive model and validation set to check the accuracy of the model (Nichols et al., 2019). In pharmacology, this ML tool can predict the effect of drug dosing for specific patient groups. For example, supervised learning was trained to predict stable warfarin dosing in Caribbean Hispanic labeled as “normal”, “low” or “high” dose patients using genetic and non-genetic clinical data (Roche-Lima et al., 2020). Unsupervised learning is used to find clusters inside a data without human labeling. In addition, subset of unsupervised learning called Deep Neural Network (DNN) could extract hidden pattern from big data. An example in pharmacology is DNN trained on transcriptomic data to classify various drugs into therapeutic categories (e.g., cardiovascular agent, antineoplastic agent, etc). From the DNN, a matrix that highlight “misclassification” of the drugs may indicates the potential for drug repurposing (Aliper et al., 2016).

The aim of this paper is to introduce a few landmark/ impactful cases of machine learning application for personalized medicine. We will identify factors that contribute to the success of machine learning in those studies. Furthermore, we try to identify threats that may obstruct the translation of machine learning into clinical practice of personalized medicine. To conclude, a concept for future development of personalized medicine is proposed.

METHODS

Google scholar search was performed with the keyword “Artificial Intelligence for Personalized Medicine” or “Machine Learning for Personalized Medicine” or combinations of these words. A few key studies were studied to showcase machine learning application in the field. First selected example is a tool called C-path that used machine learning to find novel feature from tumor images (Beck et al., 2011). This example is chosen because it highlights superiority of machine learning over human experts in novel feature extraction. However, the algorithm itself needed human expertise to guide its learning process. Second chosen example is algorithm to find “attractor metagenes” from genetics data of few types of cancers (Cheng et al., 2013). This study is chosen because it shows the potential of machine learning to find combination of genes responsible for the progression of a disease. Third chosen example is phenomapping for Heart Failure with preserved Ejection Fraction (HFpEF) (Shah et al., 2015). HFpEF phenotyping had been difficult due to heterogenous nature of the disease. This example shows the use of machine learning in unbiased patient clusterization. Fourth example is the use of ML to predict treatment outcome in Major Depressive Disorder (MDD) (Athreya et al., 2018). Fifth example is Deep Learning framework to predict response of therapy in cancer (Sakellaropoulos et al., 2019).

Important Specifications for Personalized Medicine

Before discussing specific examples, we want to propose characteristics of “good” ML model for personalized medicine. First, ML prediction must have good performance as indicated by how accurate the model could predict the outcome. Because essentially a learning tool builds a model, there is a tradeoff between training error when we are creating the model and test error when we put new data to our model. We can make a complex model with low training error by collecting more training data. However, this model will generalize poorly in the test data, a problem known as model overfitting. So, we should find good balance by making the model not too complex and not too simple (Deo, 2015).

The second characteristic is the interpretability of the

learning results. Machine learning derives important features from data by correlations. However, correlation does not imply causation (Fröhlich et al., 2018). The underlying mechanistic causation of correlation found by ML model must be investigated by human. In medicine, there is high liability when the doctor does not know why some genes are related to certain disease. A clinically unexplainable result will not likely be adopted in clinical practice.

The third characteristic is the clinical validation of the machine learning results. Thorough clinical trial must be carried out before the model found by ML can be clinically accepted. For machine learning derived hypothesis, this can be done in three steps. First is cross-validation in the initial cohort of the study. Second is external validation from another cohort to ensure there is no selection bias, for example under-represented socio-racial economic group (Gianfrancesco et al., 2018). Third validation is follow-up clinical trial to ensure the robustness of the model. The entire process is costly and time-consuming, hence only a little of ML-derived model had been clinically validated (Fröhlich et al., 2018).

To summarize, there are three important aspects to consider in machine learning for personalized medicine application. First is prediction performance, second is interpretability, third is clinical validation. Now we will explore our example with those three key points in mind.

Computational Pathologist (C-Path) (Beck et al., 2011)

Breast cancer prognostic are diagnosed by histological features of cancerous tissue. The semiquantitative epithelial features identification and grading method was first introduced by Patey and Scarff in 1928 (Patey & Scarff, 1928). Until currently, the grading process is done manually by expert pathologists. This method may suffer from inter-operator variability (Fanshawe et al., 2008). Despite newly found evidence of genes expression in cancer stroma (Bianchini et al., 2010; Finak et al., 2008), the grading of breast cancer images does not include relevant stromal features for its scoring. This highlights the subjectivity of grading method and how a novel feature finding might be biased by common practice.

Automatic classification by ML model could detect novel overlooked features and eliminate subjectivity of image reading. Koller et al developed automated algorithm called C-path to detect novel features in epithelial and stromal part of the image (Beck et al., 2011). First step of the algorithm is supervised learning to classify images as stroma or epithelial. Then, rich feature sets were extracted from both epithelial and stromal part. Lastly, images labelled as survived after 5 years and deceased after 5 years were used to build 5-Year-Survival (5YS)

predictive model. The ML algorithm found three novel stromal features that predict 5YS with good accuracy, in addition to eight epithelial features.

In term of performance, Kaplan-Meier 5YS curves of low-risk and high-risk patients were generated. The model can differentiate the two groups with statistically significant differences based on the stromal and epithelial features. In term of interpretability, the findings of stromal features correspond with biomolecular study that identifies breast cancer related stromal genes (Cheng et al., 2013). Clinical validation in this study is done by cross validation in initial cohort (Netherlands Cancer Institute, 248 patients) with eight equal folds and external validation with another cohort (Vancouver General Hospital, 328 patients) (Beck et al., 2011).

Attractor Metagenes (Cheng et al., 2013)

Different types of cancer have the same unifying capabilities such as cell invasiveness and proliferation. However, previous studies had not been successful in identifying common genes for the cancers. Cheng et al developed unsupervised clustering algorithm to find attractor metagenes from Molecular Taxonomy of Breast Cancer International Consortium data of six cancers with total genes of 11,395 (Cheng et al., 2013). The attractors found by this algorithm are strongly associated with three main processes in cancers: mesenchymal transition attractors, mitotic chromosomal instability attractors, and lymphocyte-specific attractors.

In term of performance, Kaplan-Meier survival curves of high metagene and low metagene level were plotted. From the plot, the mesenchymal transition metagenes are found to yield statistically significant result in early-stage breast cancer data. Mitotic metagene attractors could discriminate two patients' group from breast cancer data. Lymphocyte metagene affecting few nodes could distinguish survival rate between the high genes and low genes group. In term of interpretability, it should be noted that the attractor metagenes could already be associated with specific phenotypes such as invasiveness of the cell and uncontrollable cell division. However, further biomolecular study still needs to be performed to find the mechanics behind the genes and the effects. Lastly, the data came from database of International Consortium, and validation with external cohort was not performed.

Phenomapping of Heart Failure with preserved Ejection Fraction (HFpEF) (Shah et al., 2015)

HFpEF is a cardiac disorder with heterogeneous phenotype. To map HFpEF patient into distinct phenotype groups (i.e. phenomapping), Shah et al developed unsupervised machine learning algorithm to do the task. First, Bayesian Information Criterion Analysis is used to

find the optimum number of distinct phenotype groups. Then, the algorithm sort detailed clinical, laboratory, electrocardiographic, and echocardiographic data from 397 participants to distinct three phenotypes group (Shah et al., 2015). Subsequently, different phenogroup is associated risk data by Kaplan Meier Curve, and it is found that they have different risk factor. The authors claimed that their study is the first one to use high-density phenomapping of a cardiovascular syndrome. Furthermore, stratifying patients into different phenotype group is beneficial for personalized therapies.

From performance viewpoint, Kaplan-Meier curve showed that the three phenotypes' groups were distinct. In term of interpretability, they have identified the three groups as: (1) younger patients with normal B-type Natriuretic Peptide (BNP) and moderate diastolic dysfunction, (2) obese and diabetic patients with obstructive sleep apnea and worse Left Ventricle relaxation, (3) older patients with significant kidney illness, electrical and myocardial remodeling, pulmonary high blood pressure, and Right Ventricle disfunction. Interpretation of the data were concluded from examining relevant features in each group. The model developed was validated with prospectively enrolled cohort of 107 patients (Shah et al., 2015).

Machine Learning (ML) to Predict Treatment Outcome in Major Depressive Disorder (MDD) (Athreya et al., 2018)

In diseases that have complex phenotypes such as Major Depressive Disorder (MDD), the subject/ doctor-rated severity of symptoms is poor predictor of therapeutic success (Chekroud et al., 2016). The outcome prediction does not include the biological measures that reflect the molecular mechanism of drugs action, and by only clinical measures the performance was slightly better than chance. Athreya et al developed ML model that could predict the outcome of antidepressant treatment in patients with MDD by using combination of metabolomics, genomics, and sociodemographic factors. The developed model was constructed from data of the Mayo Clinic Pharmacogenomics Research Network Antidepressant Medical Pharmacogenomic Study with 603 patients completed the trial (Holmes et al., 2011). The patients were given citalopram for eight weeks. The psychiatric assessment of depression severity was collected at baseline, four weeks, and eight weeks. Seven million single-nucleotide polymorphisms (SNPs) and 31 metabolites were taken from 290 of the 603 patients at baseline, four weeks, and eight weeks interval (Athreya et al., 2018). A model was built by biological data, clinical measure, and sociodemographic factors as predictor variables. Except *seasonal pattern* and *involvement item*, all top five predictors in men and women were related to metabolomics. Furthermore, the

metabolites identified were known to be associated with mood in behavioral sciences. In overall, the model which includes omics and sociodemographic factors increased the prediction accuracy from 35% to 80% compared with using the clinical measures alone (Athreya et al., 2018).

Deep Learning Framework for Predicting Response to Therapy in Cancer (Sakellaropoulos et al., 2019)

Sakellaropoulos et al trained Deep Neural network (DNN) from dataset of 1001 cell lines and 251 drugs from the Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang et al., 2013). Its performance was then evaluated on patient cohort from 4 datasets: (1) The Cancer Genome Atlas (TCGA) for two drugs: cisplatin and paclitaxel (Ding et al., 2016), (2) unpublished data from clinical trial of PARP inhibitor at MD Anderson Cancer Center, (3) a cohort of esophageal adenocarcinomas treated with neoadjuvant chemotherapy (Beckell et al., 2019), and (4) phase II/III clinical trial of bortezomib in patients with relapsed multiple myeloma (Mulligan et al., 2007). The performance of DNN is evaluated with two classical ML method: Random Forest (RF) (Costello et al., 2014) and elastic net (Enet) (Zou & Hastie, 2005). Overall, DNN performed better than RF and Enet as evaluated by effect size and statistical Wilcoxon test. From AUC (area under the curve), DNN scored 65%, better than RF (54%) and Enet (52%) (Sakellaropoulos et al., 2019, figure 3C). To verify biological mechanism behind the DNN finding, they evaluated the first hidden layer of the pathways detected by DNN for the following drugs: bortezomib, cisplatin, and paclitaxel. Evaluation with prior literature gave high correspondence: 79% for bortezomib, 96% for cisplatin, and 68% for paclitaxel (Sakellaropoulos et al., 2019). This means that DNN may be effective in revealing underlying pathway of pharmacogenomics response to drugs in cancer treatment.

Evaluation of The Machine Learning Models

Table 1 outlines the evaluation of reviewed systems with parameters of performance, interpretability, and clinical validation.. High performance means the algorithm achieved high predictive accuracy between different subgroups. High interpretability means biomolecular causation can be related with the model derived from ML algorithm. Medium interpretability means some causation can be associated with the ML model. Low interpretability means no association can be made between ML correlation and underlying biological factors. Medium clinical validity means the algorithm was tested with internal and external cohort. Low clinical validity means the algorithm was tested only within internal cohort.

C-path findings were propelled by biomolecular hints that stromal genes is important for the prognostic of

Table 1. Evaluation of the reviewed systems based on their performance, interpretability, and clinical validity

Criterion	C-Path	Attractor Metagenes	Phenomapping of HFpEF	Antidepressant Effect Predictors in MDD	DNN for Predicting Response of Therapy in Cancer
Performance	High	High	High	High	High
Interpretability	High	Medium	Low	Medium	High
Clinical Validity	Medium	Low	Low	Low	Medium

breast cancer (Bianchini et al., 2010). Therefore, the ML algorithm in this case is hypothesis driven. The first part of the learning algorithm was supervised learning with stromal and epithelial part of the image annotated by expert (Beck et al., 2011). After that, unsupervised learning learns new features from the stromal and epithelial part. These features can distinguish risk groups in patients, resulting in high performance score. Stromal features identified by machine learning algorithm corresponds strongly to previous study of biomolecular stromal markers, therefore this study had high clinical interpretability. Nonetheless, clinical validation with larger cohort needs to be done for the wide application of this finding. In attractor metagenes study, the genes found can be associated with distinct events in cancer prognosis. However, the causation from genes to the specific events need further investigation. Therefore, its interpretability score was medium. Kaplan Meier curve shows that distinct patient group can be identified, but since the data used are from database and not clinical trial, its clinical validity is low. Further testing with large cohort is needed to ensure validity of the model. Phenomapping of HFpEF shows three distinct patient groups, therefore it has high performance score. Although different characteristic of patient groups can be identified, it is not clear how different patient group corresponds to different disease type, so we concluded that its interpretability is low. Since validation was performed only from internal cohort, the clinical validity is low. In this study, clinical explanation and further validation is needed for the clinical adoption of the algorithm.

These first three examples showed application of AI to extract information from data and to stratify patient groups. First example used images data, second example used gene profile data, and third example used health record data. This showed that various modalities can be used for personalized medicine. Further studies to investigate how different modalities could be used for AI patient stratification need to be done. For example, imaging and health record data could be strong predictors to stratify risk in steatohepatitis patient.

In the ML model build for predicting antidepressant effect

(Nguyen et al., 2021), its accuracy reached 80%, which we considered as high. The metabolomics predictors were associated with mood, but clear mechanism of how this metabolomics interact with drug is unexplored, so the interpretability is medium. The validation done was only within initial cohort, so its clinical validity is low. The DNN model for predicting effect of drugs in cancer achieved high performance (Sakellaropoulos et al., 2019). Furthermore, the drug pathways identified by DNN corresponds with previous literature, so the interpretability is high. The DNN is validated on 4 cohorts, so the clinical validation score is medium. This last example showed oncology as the most promising and actively researched field for personalized medicine (Tsimberidou, Fountzilas, Nikanjam, & Kurzrock, 2020).

The availability of representative data is crucial for the AI model. The study of attractor genes have benefitted from data available from consortium (Cheng et al., 2013). Longitudinal study could provide large and representative data (Beck et al., 2011; Shah et al., 2015). Data from cohorts in previous trial can be extracted for building ML model (Nguyen et al., 2021; Sakellaropoulos et al., 2019). In addition, multimodalities data could achieve better prediction for drug response, as demonstrated by drug response prediction in Major Depressed Disorder study (Nguyen et al., 2021). Collection and standardization national or international bank of multimodalities health-related data (genomics, biomarkers, images, phenotype, sociodemographic, etc) is crucial for the research communities to extract AI-driven hypothesis for various diseases and drugs. National and international collaboration backed with government initiative should be carried out for nationwide data management, such as UK Biobank initiative (Sudlow et al., 2015).

Human guidance can teach the ML model to improve its performance. This is showed in first example (C-Path) by how hypothesis is used by the expert to annotate the stromal part, and this hypothesis is then proven by AI method finding. In contrary, ML model can also guide human to find new features/predictors. Attractor genes identified by AI method in second example should be studied further to elucidate why it has correlation with different phenotype. Patient classification results in

third example can be further studied to find relationship between disease subtype and to find why a group develop such subtype. Both human and ML model can “learn” from each other.

As last note, we saw that most AI finding are retrospective, meaning the data was extracted from previous clinical trials in other studies. No research is actively validating the finding of ML model in clinical validation. More progressive uptake of ML model by biological study, followed by clinical trial is important for proving the usefulness of such ML model for personalized medicine.

CONCLUSION

With the explosion of genes sequencing, imaging, bioelectronic health record, and other medical data there is a need to extract important information for designing therapy as specific as possible which depends on patient individual characteristic, also known as “personalized medicine/ precision medicine”. Achieving personalized treatment would be impossible without AI/ML tools to build model from those data. To develop good AI model, first we need to collect and standardize representative and inclusive data. In addition, multi-modalities data could enhance the performance of AI model. After getting standardized multimodal data in well-managed national/international databanks, experts or disease specialist should guide the application of AI toward personalized medicine of specific disease. Then, finding from AI model must be investigated by the clinical community to understand the underlying causation. Finally, after the mechanism is understood, the AI prediction model must be validated by larger clinical trial for its translation into clinical practice.

REFERENCES

Abul-husn, N. S., & Kenny, E. E. (2019). Perspective Personalized Medicine and the Power of Electronic Health Records. *Cell*, 177(1), 58–69. <https://doi.org/10.1016/j.cell.2019.02.039>

Agerbo, E., Sullivan, P. F., Vilhjalmsson, B. J., & Pedersen, C. B. (2015). Polygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the Risk for Scizophrenia: A Danish Population-Based Study and Meta Analysis. *Jama Psychiatry*, 72, 635–641.

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., & Zhavoronkov, A. (2016). Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular Pharmaceutics*, 13(7), 2524–2530. <https://doi.org/10.1021/acs.molpharmaceut.6b00248>

Athreya, A., Iyer, R., Neavin, D., Wang, L., Weinshilboum, R., Kaddurah-Daouk, R., ... Bobo, W. (2018). Augmentation of physician assessments with multi-omics enhances predictability of drug response: A case study of major depressive disorder. *IEEE Computational Intelligence Magazine*, 13(3), 20–31. <https://doi.org/10.1109/MCI.2018.2840660>

Beck, A. H., van de Rijn, M., van de Vijver, M. J., Sangoi, A. R., Koller, D., Marinelli, R. J., ... Nielsen, T. O. (2011). Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science Translational Medicine*, 3(108), 108ra113-108ra113. Retrieved from <http://stm.sciencemag.org/cgi/doi/10.1126/scitranslmed.3002564%5Cnfile:///Files/34/34AF3276-62E1-4587-B757-ECB6DA0E3FB9.pdf%5Cnpapers3://publication/doi/10.1126/scitranslmed.3002564>

Bianchini, G., Qi, Y., Alvarez, R. H., Iwamoto, T., Coutant, C., Ibrahim, N. K., ... Pusztai, L. (2010). Molecular Anatomy of Breast Cancer Stroma and Its Prognostic Value in Estrogen Receptor – Positive and – Negative Cancers. *Journal of Clinical Oncology*, 28, 4316–4323. <https://doi.org/10.1200/JCO.2009.27.2419>

Blackstone, E. H. (2019). Precision Medicine Versus Evidence-Based Medicine: Individual Treatment Effect Versus Average Treatment Effect. *Circulation*, 140(15), 1236–1238. <https://doi.org/10.1161/CIRCULATIONAHA.119.043014>

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., ... Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, 3(3), 243–250. [https://doi.org/10.1016/S2215-0366\(15\)00471-X](https://doi.org/10.1016/S2215-0366(15)00471-X)

Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F., & Li, L. (2011). China Kadoorie Biobank of 0 . 5 million people : survey methods , baseline characteristics and long-term follow-up. *International Journal of Epidemiology*, 40(September), 1652–1666. <https://doi.org/10.1093/ije/dyr120>

Cheng, W., Yang, T. O., & Anastassiou, D. (2013). Biomolecular Events in Cancer Revealed by Attractor Metagenes. *Plos Computational Biology*, 9(2). <https://doi.org/10.1371/journal.pcbi.1002920>

Clayton, T. A., Lindon, J. C., Cloarec, O., Antti, H., Charuel, C., Hanton, G., ... Nicholson, J. K. (2006). Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature*, 440(7087), 1073–1077. <https://doi.org/10.1038/nature04648>

- Collins, R. (2012). What Makes UK Biobank special? *Lancet*, 379, 1173–1174.
- Costello, J. C., Heiser, L. M., Georgii, E., Gönen, M., Menden, M. P., Wang, N. J., ... Van Westen, G. J. P. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. *Nature Biotechnology*, 32(12), 1202–1212. <https://doi.org/10.1038/nbt.2877>
- Deo, R. C. (2015). Machine Learning in Medicine. *Circulation*, 132(20), 1920–1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>. Machine
- Ding, Z., Zu, S., & Gu, J. (2016). Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics*, 32(19), 2891–2895. <https://doi.org/10.1093/bioinformatics/btw344>
- Fanshawe, T. R., Lynch, A. G., Ellis, I. O., Green, A. R., & Hanka, R. (2008). Assessing Agreement between Multiple Raters with Missing Rating Information , Applied to Breast Cancer Tumour Grading. *PLoS ONE*, 3(8). <https://doi.org/10.1371/journal.pone.0002925>
- Finak, G., Bertos, N., Pepin, F., Sadekova, S., Souleimanova, M., Zhao, H., ... Park, M. (2008). Stromal gene expression predicts clinical outcome in breast cancer. *Nature Medicine*, 14(April), 518–527. <https://doi.org/10.1038/nm1764>
- Frankell, A. M., Jammula, S. G., Li, X., Contino, G., Killcoyne, S., Abbas, S., ... Fitzgerald, R. C. (2019). The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nature Genetics*, 51(3), 506–516. <https://doi.org/10.1038/s41588-018-0331-5>
- Fröhlich, H., Balling, R., Beerenwinkel, N., Kohlbacher, O., Kumar, S., Lengauer, T., ... Zupan, B. (2018). From hype to reality: data science enabling personalized medicine. *BMC Medicine*, 16(1), 150. <https://doi.org/10.1186/s12916-018-1122-7>
- Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine*, 178(11), 1544–1547. <https://doi.org/10.1001/jamainternmed.2018.3763>. Potential
- Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W. A., Li, R., Manolio, T. A., ... Network, T. (2013). Open The Electronic Medical Records and Genomics (eMERGE) Network : past, present, and future. *Genetics in Medicine*, 15(10), 761–771. <https://doi.org/10.1038/gim.2013.72>
- Hoffman, M. A., & Williams, M. S. (2011). Electronic medical records and personalized medicine. *Human Genetics*, 130, 33–39.
- Holmes, M. V, Perel, P., Shah, T., Hingorani, A. D., Casa, J. P., Hoskins, J. M., ... Weinshilboum, R. M. (2011). Use of omeprazole as a probe drug for CYP2C19 phenotype in Swedish Caucasians: comparison with S-mephenytoin hydroxylation phenotype and CYP2C19 genotype. *Pharmacogenetics and Genomics*, 5(6), 335–398.
- Huopaniemi, I., Nadkarni, G., Nadukuru, R., Lotay, V., Ellis, S., Gottesman, O., & Bottinger, E. P. (2014). Disease progression subtype discovery from longitudinal EMR data with a majority of missing values and unknown initial time points. In *AMIA Annual Symposium Proceedings* (pp. 709–718).
- Ingelsson, E., & McCarthy, M. I. (2018). Human Genetics of Obesity and Type 2 Diabetes Mellitus. *Circulation: Genomic and Precision Medicine*, 11(June), 1–12. <https://doi.org/10.1161/CIRCGEN.118.002090>
- Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 13, 395–405.
- Kensler, T. W., Spira, A., Garber, J. E., Szabo, E., Lee, J. J., Dong, Z., ... Lippman, S. M. (2016). Transforming cancer prevention through precision medicine and immune-oncology. *Cancer Prevention Research*, 9(1), 2–10. <https://doi.org/10.1158/1940-6207.CAPR-15-0406>
- Mak, A. C. Y., White, M. J., Eckalbar, W. L., Szpiech, Z. A., Oh, S. S., Pino-yanes, M., ... Hernandez, R. D. (2018). Whole-Genome Sequencing of Pharmacogenetic Drug Response in Racially Diverse Children with Asthma. *American Journal of Respiratory and Critical Care Medicine*, 197(12), 1552–1564. <https://doi.org/10.1164/rccm.201712-2529OC>
- Mulligan, G., Mitsiades, C., Bryant, B., Zhan, F., Chng, W. J., Roels, S., ... Anderson, K. C. (2007). Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood*, 109(8), 3177–3188. <https://doi.org/10.1182/blood-2006-09-044974>
- Nagai, A., Hirata, M., Kamatani, Y., Muto, K., & Matsuda, K. (2017). Overview of the BioBank Japan Project : Study design and profile. *Journal of Epidemiology*, 27, S2–S8. <https://doi.org/10.1016/j.je.2016.12.005>
- Nguyen, T. T. L., Liu, D., Ho, M. F., Athreya, A. P., &

- Weinshilboum, R. (2021). Selective Serotonin Reuptake Inhibitor Pharmacogenomics: Mechanisms and Prediction. *Frontiers in Pharmacology*, 11(January), 1–10. <https://doi.org/10.3389/fphar.2020.614048>
- Nichols, J. A., Herbert Chan, H. W., & Baker, M. A. B. (2019). Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*, 11(1), 111–118. <https://doi.org/10.1007/s12551-018-0449-9>
- Patey, D. H., & Scarff, R. W. (1928). The position of histology in the prognosis of carcinoma of the breast. *Lancet*, 801–804.
- Roche-Lima, A., Roman-Santiago, A., Feliu-Maldonado, R., Rodriguez-Maldonado, J., Nieves-Rodriguez, B. G., Carrasquillo-Carrion, K., ... Duconge, J. (2020). Machine learning algorithm for predicting warfarin dose in caribbean hispanics using pharmacogenetic data. *Frontiers in Pharmacology*, 10(January), 1–8. <https://doi.org/10.3389/fphar.2019.01550>
- Sakellaropoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., ... Gorgoulis, V. G. (2019). A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Reports*, 29(11), 3367–3373.e4. <https://doi.org/10.1016/j.celrep.2019.11.017>
- Shah, S. J., Katz, D. H., Selvaraj, S., Burke, M. A., Yancy, C. W., Gheorghide, M., & Bonow, R. O. (2015). Phenomapping for Novel Classification of Heart Failure with Preserved Ejection Fraction. *Circulation*, 131(3), 269–279. <https://doi.org/10.1161/CIRCULATIONAHA.114.010637.Phenomapping>
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26, 1135–1145.
- Singer, J., Irmisch, A., Ruscheweyh, H. J., Singer, F., Toussaint, N. C., Levesque, M. P., ... Beerenwinkel, N. (2017). Bioinformatics for precision oncology. *Briefings in Bioinformatics*, 20(3), 778–788. <https://doi.org/10.1093/bib/bbx143>
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., ... Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine*, 12(3), 1–10. <https://doi.org/10.1371/journal.pmed.1001779>
- Tsimberidou, A. M., Fountzilas, E., Nikanjam, M., & Kurzrock, R. (2020). Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer Treatment Reviews*. <https://doi.org/10.1016/j.ctrv.2020.102019>
- Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., ... Garnett, M. J. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(D1), 955–961. <https://doi.org/10.1093/nar/gks1111>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/https://doi.org/10.1111/j.1467-9868.2005.00503.x>