An in-silico approach for drug repositioning to tumour anti-migration using an integrated genomic strategy

Yong Mao[1*], Kemi Cui[1*], Wang Lulu[1,2], Hong Zhao[1], Fang Nie[1,3],
Miriam Brandl[1,4], Dominik Beck[1,4], Liang Gao[1,5], Stephen Wong[1,2]

1. Bioengineering and Bioinformatics Program, The Methodist Hospital Research Institute, Weill Cornell Medical College, Houston, Texas 77030, USA
2. School of Medicine, Shandong University, Jinan, Shandong 250012, China
3. Jiangsu Key Laboratory of Molecular Imaging and Functional Imaging, Department of Radiology, Zhongda Hospital, Southeast University, Nanjing, Jiangsu, 210009, China
4. School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT, 2600, Australia
5. Department of Bioengineering, Rice University, Houston, Texas, 77005, USA

*. The authors contributed equally to this work.

***Abstract*— Cell migration is a key step for deterioration of many *in situ* or metastasis malignant tumours. Tumour anti-migration is a promising strategy to treat cancer, but corresponding drugs developed under such a strategy are still in dire poverty, partly due to the lengthly process of drug trials and approval required by the US Food and Drug Administration (FDA). Given there are thousands of FDA approved drugs in the market, we believe that drug repositioning may provide a fast and cost-effective way to identify potential anti-migration drugs. In this paper, an in-silico drug screening method using a genomic strategy is proposed for the goal, in which genomic signature identification combined with support vector machine modelling is adopted to estimate drug efficacy. And a high-throughput, sensitive, 3-dimensional invasion assay by quantitative bioluminescence imaging proved the performance of proposed method on *in vitro* disease models.**

I. INTRODUCTION

Although cancer usually starts as a localized disease, at its late-stage, a pernicious progression is the invasion of tumour cells into surrounding tissues, and even resulting in the development of distant metastases, which is a major cause of death. Active migration of tumour cells is a prerequisite for tumour-cell invasion and metastasis, thus a strategy to stop tumour progression is to develop drugs that regulate or inhibit the migratory activity of malignant tumour cells [1]. To be effective, usually drug combinations are required for suppression of cancer migration, however, very few effective anti-migration drugs exist in the market.

The current paradigm of drug discovery fails to tackle disease with complex mechanisms. Drug repurposing or reposition, referring to the identification and development of new uses for existing drugs, has recently shown promise as a cost-effective and faster way to deliver cure to the patients, and costs much less than de novo drug discovery and development [2] due to their known pharmacokinetics and safety profiles, and any newly identified drug reuse can be rapidly evaluated in phase II clinical trials.

Due to the complex nature of disease mechanism and drug's mode of action, screening thousands of available drugs using specific disease models in the laboratory is time-consuming and costly. Recently, more and more disease- and drug-related experimental data are becoming public available; our hypothesis is that we may be able to make use of these datasets to design a virtual screening engine for drug repositioning. A dozen of in-silico approaches have recently been proposed to discover the connection between disease and drug compounds, however, rare of them are well-recognized globally for combining disease mechanism discovery with drug repositioning, most are still focused on 'me-too' drug candidate identification.

Gene expression profiles are used extensively in disease diagnosis, prognosis, and prediction of the response to therapies. Since gene expression signatures could be used to describe the mechanism of disease and how therapeutics will work on the disease pathway, they can serve as bridges to connect diseases and drugs together. In gene expression signatures, genes are usually organized in modules, in which different genes act in concert to carry out a specific function. Disease mechanism may be highly related to behavior of a number of modules, but in reality, possibly only some sub-parts of a subgroup of these modules are relevant to disease mechanism as well as patients' response to therapeutics [3].

In this paper, a new in-silico drug reposition method using a genomic strategy is proposed on screening drug compounds for tumour anti-migration. Experimental results achieved by a 3-dimensional cell invasion assay using quantitative bioluminescence imaging proved the effectiveness of the proposed method. Section 2 describes details of the strategy. Two cases for triple-negative breast cancer (TNBC) and glioblastoma multiforme (GBM) are analysed accordingly in Section 3 while Section 4 provides the conclusion.

II. VIRTUAL DRUG SCREENING USING A GENOMIC STRATEGY

978-1-4577-0422-2/11/$26.00 ©2011 IEEE

From a genomic view, the core mechanism of cancer is highly related to a group of genes, usually called gene signature, which could be used to link the effect of drugs with diseases as bridges. But genes significantly different between disease and controls are not all necessarily relevant to the underlying disease mechanism. So the core signature extraction is a key step in our method as stage 1. By evaluating whether and how a drug's pattern of mode of actions (MOA) changes the pattern of disease behaviour reflected on the signature, an in-silico inference could be made for effective drug screening. In stage 2, a support vector machine modeling method is proposed to implement this inference. Finally, *in vitro* laboratory validation is used to evaluate the effectiveness of the virtual screening method.

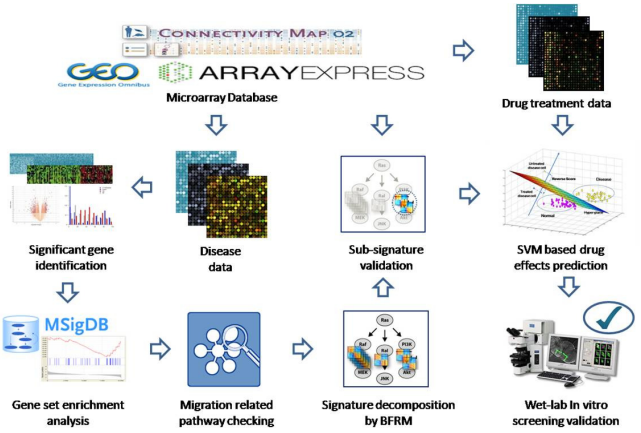

Fig. 1 Workflow chart of the virtual screening method for drug repositioning.

Details for the method are shown in Figure 1. First, gene microarray data for a specific disease and candidate drugs are collected from public available sources, e.g. ArrayExpress and GEO. Based on clinical and *in vitro* disease gene expression data, a pre-signature is identified by fold change and T-test, and then significant parts in the pre-signature are annotated by canonical pathway enrichment analysis referred to MSigDB, by which disease mechanism related part could be reflected. Then the top tumour migration related pathways are indicated according to the knowledge extracted from Ingenuity database. Combining these chosen pathways together as the new signature for Bayesian factor regression modelling (BFRM) method [3], it could be decomposed into a dozen of sub-signatures, of which those highly related to clinic outcomes are united together as a refined signature to describe key mechanism for the disease.

A quantitative model to describe how disease is different from normal is built by support vector machine (SVM) algorithm using the extracted signature, whose principle is shown in Figure 2. In this model, a transition direction to describe the trend from normal to disease is extracted, which is perpendicular to the hyper-plane solved by SVM. We hypothesize that if a drug candidate could change the status of a disease more along the direction reversely, it will have a higher chance to cure the disease. Accordingly, a reverse score (RS) is defined to quantify or predict drug effects by projecting their related microarray data onto the direction. RS is calculated by the difference between the disease objects with or without drug treatment. The relative performance of all drugs could be compared or ranked according to their RSs.

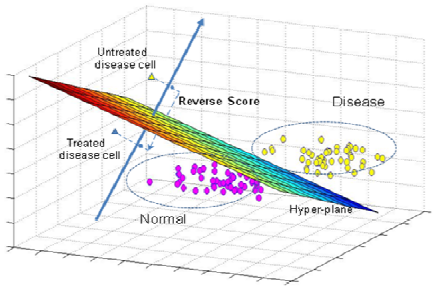

Fig. 2 Drug effect prediction model by support vector machine. Yellow and blue circles denote disease and normal samples without treatment respectively; the hyper-plane is produced by SVM, and the blue solid arrow means the transition from normal to disease; the yellow triangle shows the disease object's condition before drug treatment, the purple one illustrates the status of disease object after drug treatment, the vector projected on the extracted direction between the two triangles denotes the Reverse Score (RS), which means to what extent the specific drug could reverse the disease.

After all candidate drugs are scored and ranked, a high-throughput 3-dimensional cell invasion (BLI) assay using quantitative bioluminescence imaging [4] is employed to test the anti-migration performance of the drug compounds on *in vitro* disease model. The basic principal of BLI assay could be described as follows: in each well of the PDL plate, disease cells expressing luciferase are set on top of a 3D cellular matrix, which fills the whole well, at the beginning of experiment. Then, the cancer cells usually will migrate to the well bottom and will react with fluorescent substance, D-luciferin, anchored there, and the fluorescence produced will be imaged to quantify its migration ability. If a strong anti-migration drug is added in, little fluorescence will be detected. Meanwhile, a parallel experiment is set by combining D-luciferin with disease cells from the beginning, so drug effects on cellular proliferation could be estimated, assuming it's an important influence factor for evaluating anti-migration functions of the drug. Considering them together, an index called invasion ratio is calculated to evaluate the overall anti-migration effects. The experimental flow chart is shown in Figure 3. In our work, each drug is tested under 3 titration doses, 1, 10 and 100 μm. Experimental quality control is implemented for all drugs according to their reproducibility. For evaluating our prediction results more efficiently, we assume, if the Invasion ratio of a drug is between 80% and 120% compared with the control, it is considered with no effect. If it's above 120%, it will promote cancer cell's migration ability. Otherwise, it inhibits cell's migration. Assuming anti-migration effect as -1, no effect as 1, and pro-migration effect as 1, each drug's effects under the 3 dosages are added together to form an experimental score (ES) to describe its anti-migration ability.

Finally, Pearson correlation coefficient is used to evaluate whether our drug prediction results are corresponding to their anti-migration ESs in a global view. The confusion matrix is

used to evaluate drug screening accuracy based on drugs' real anti-migration and pro-migration effects. And the top ranked drugs' mode of action will be estimated according to their influence on the extracted signature.

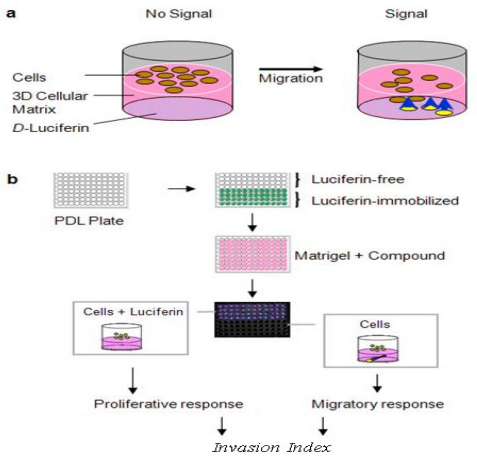

Fig. 3 Description of BLI 3D assay platform: a. basic principal of BLI experiments, b. BLI 3D assay for high-throughput anti-migration drug screening.

III. RESULTS

Two cases from triple negative breast cancer (TNBC) and glioblastoma multiforme (GBM) respectively are tested using the proposed method.

A. *Anti-migration drug screening for TNBC*

TNBC is a particular aggressive subtype, which has a much higher metastasis rate than other breast cancer subtypes. The median survival of untreated patients or those treated with corticosteroids alone is about one month and unfortunately no effective therapeutic regimens exist.

Using clinical microarray data provided in [5], 88 pathways are identified significantly according to gene sets enrichment analysis (GSEA), and by *in vitro* data in [6], 77 pathways are identified. Across these two lists, there are 30 pathways in common, in which 5 pathways are maintained for their possible relation to tumour migration, according to their function description as well as their significant p values. These five pathways are 'Breast cancer estrogen signalling,' 'Cell communication,' 'P53 signaling pathway,' 'Focal adhesion,' and 'ECM receptor interaction.'

The new signature formed by all genes involved in these five pathways is decomposed into 15 sub-modules, in which 3 sub-modules are maintained according to their relationship with patients' survival time by Kaplan-Meier survival analysis, and their relationship with tumour metastasis status by T-test, as well as their repeatability on similar clinical validation dataset from [7].

Forty-five genes from these 3 sub-modules are maintained as features to build a support vector machine model. In order to enlarge the training sample size, all clinical datasets used are combined together, where COMBAT algorithm [8] is employed to reduce the batch effects among all data. According to the availability of the drugs in our lab and their corresponding data in the Connectivity map [9], thirty-eight drugs are chosen to do *in vitro* biological validation, and 31 of them are maintained across quality control for this case.

SVM model is trained based on normalized clinical data. Then drug related data is projected on the model to calculate corresponding RSs. For each drug, whose corresponding microarrays are produced from different cell lines as well as different dosages, all related RSs are averaged to form a comprehensive score to predict its tumour anti-migration effect. If the score of a drug is greater than 0, it is predicted as an anti-migration (PAM) drug, otherwise it's predicted as pro-migration (PPM).

The PCC between predicted scores and that from experiment is 0.48, which is better than the PCC value, 0.13, achieved based on the signature before using BFRM. The result achieved by incorporating BFRM is significant different from random results by P=0.0031. By the comparison, it means that the signature refinement is necessary and important for our method. Another result for our method is shown in Table I by the confusion matrix, where all experimental results are categorized according to their ESs, $ES \geqslant 2$ is set as with anti-migration (AM) effect, $-1 \leqslant ES \leqslant 1$ as no effect (NE), and $ES \leqslant -2$ as pro-migration (PM) effect.

TABLE I
CONFUSION MATRIX FOR DRUG PREDICTION TO TNBC

Prediction	**Wet-lab *in vitro* experimental results**		
	AM	**NE**	**PM**
PAM	4	8	2
PPM	1	10	6

AM: anti-migration, NE: no effect, PM: pro-migration, PAM: predicted anti-migration, PPM: predicted pro-migration.

According to our prediction results, the most significant tumour anti-migration and pro-migration drugs are Etodolac and Mifepristone respectively, which are consistent with that observed in our *in vitro* screening experiments. Etodolac is an approved nonsteroidal anti-inflammatory agent and has been reported with cell migration and adhesion inhibition ability in bladder cancer [10]. Mifepristone is an approved drug for treating hypercortisolism in patients with nonpituitary cushing syndrome. It has been reported with pro-migration effect in prostate cancer cell invasion by influencing cells' focal adhesion function [11]. By analysing their influence on the refined signature, it is found that Etodolac has significant effect on the 'Breast cancer estrogen signalling' part and Mifepristone has influence on 'Focal adhesion and ECM receptor interaction' part.

B. *Anti-migration drug screening for GBM*

Glioblastoma multiforme (GBM) is the most biologically aggressive subtype in malignant gliomas. GBM tumour cells persistently migrate to or infiltrate adjacent surrounding tissues, and cause a higher lethal rate as well as shorter survival time for the patients. It is necessary to develop effective drugs to control GBM, especially when drug resistance happens to current available drugs.

118 pathways are identified significant by clinical data from [12], and 102 pathways are identified by *in vitro* data from GSE12305. There exist 24 mutual pathways. According to pathways' description and their p values, 6 pathways are kept, which are 'Glioma,' 'MAPK signaling pathway,' 'P53 signaling pathway,' 'Focal Adhesion,' 'ECM receptor interaction,' and 'Regulation of the actin cytoskeleton by Rho GTPases.'

The signature formed by all genes from these six pathways is further decomposed into 22 sub-modules. Three sub-modules are chosen according to their relationship with patients' survival time and their repeatability on clinical validation set from [13, 14].

Using thirty-six genes from these 3 sub-modules, a SVM model is built. Thirty drugs from 38 drugs are maintained according to experimental quality control, and their corresponding data is projected on the model. The PCC between predicted scores and experimental scores is 0.52, which outperforms the PCC value, 0.16, achieved based on the signature without using BFRM. The result achieved by incorporating BFRM is significant different from random results by P=0.0017. These proved again the fact that a refined signature is important for such approach. The prediction performance by confusion matrix is shown in TABLE II.

TABLE III
CONFUSION MATRIX FOR DRUG PREDICTION TO GBM

Prediction	**Wet-lab *in vitro* experimental results**		
	AM	**NE**	**PM**
PAM	2	6	3
PPM	1	10	8

AM: anti-migration, NE: no effect, PM: pro-migration, PAM: predicted anti-migration, PPM: predicted pro-migration.

According to prediction results, the most significant anti-migration and pro-migration drugs are Isotretinoin and Isoprenaline respectively. In our wet-lab *in vitro* screening experiments, the consistent results are observed. Isotretinoin is a medication approved for severe acne and certain skin cancers. It has been reported with inhibition function for neural crest cell [15]. Isoprenaline is used mainly as bronchodilator and heart stimulant. By analysing their influence on former refined signature for GBM, Isotretinoin has significant effect on the 'MAPK Signaling pathway' part and Isoprenaline will influence the whole signature comprehensively.

IV. CONCLUSIONS

In this work, an in-silico drug screening method by an integrated genomic strategy is proposed for drug reposition, where genomic signature identification and support vector machine modelling are combined to estimate drugs' efficacy. The performance of proposed method is proved using a sensitive 3-dimensional invasion assay by quantitative bioluminescence imaging on *in vitro* disease models. Meanwhile, the model could be used to estimate the drug's mechanism partially.

ACKNOWLEDGMENT

This project is supported by John S Dunn distinguished endowed chair funding to STCW.

REFERENCES

[1] Entschladen F, Drell TLt, Lang K, Joseph J, Zaenker KS. Tumour-cell migration, invasion, and metastasis: navigation by neurotransmitters. The lancet oncology 2004;5(4):254-8.

[2] Boguski MS, Mandl KD, Sukhatme VP. Drug discovery. Repurposing with a difference. Science (New York, NY 2009;324(5933):1394-5.

[3] Chang JT, Carvalho C, Mori S, Bild AH, Gatza ML, Wang Q, Lucas JE, Potti A, Febbo PG, West M, Nevins JR. A genomic strategy to elucidate modules of oncogenic pathway signaling networks. Molecular cell 2009;34(1):104-14.

[4] Zhao H, Tang C, Cui K, Ang BT, Wong ST. A screening platform for glioma growth and invasion using bioluminescence imaging. Laboratory investigation. Journal of neurosurgery 2009;111(2):238-46.

[5] Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S. X chromosomal abnormalities in basal-like human breast cancer. Cancer cell 2006;9(2):121-32.

[6] Novak P, Jensen T, Oshiro MM, Wozniak RJ, Nouzova M, Watts GS, Klimecki WT, Kim C, Futscher BW. Epigenetic inactivation of the HOXA gene cluster in breast cancer. Cancer research 2006;66(22):10664-70.

[7] Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM. Concordance among gene-expression-based predictors for breast cancer. The New England journal of medicine 2006;355(6):560-9.

[8] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England) 2007;8(1):118-27.

[9] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science (New York, NY 2006;313(5795):1929-35.

[10] Okamoto A, Shirakawa T, Bito T, Shigemura K, Hamada K, Gotoh A, Fujisawa M, Kawabata M. Etodolac, a selective cyclooxygenase-2 inhibitor, induces upregulation of E-cadherin and has antitumour effect on human bladder cancer cells in vitro and in vivo. Urology 2008;71(1):156-60.

[11] Yan J, Erdem H, Li R, Cai Y, Ayala G, Ittmann M, Yu-Lee LY, Tsai SY, Tsai MJ. Steroid receptor coactivator-3/AIB1 promotes cell migration and invasiveness through focal adhesion turnover and matrix metalloproteinase expression. Cancer research 2008;68(13):5460-8.

[12] Hodgson JG, Yeh RF, Ray A, Wang NJ, Smirnov I, Yu M, Hariono S, Silber J, Feiler HS, Gray JW, Spellman PT, Vandenberg SR, Berger MS, James CD. Comparative analyses of gene copy number and mRNA expression in GBM tumours and GBM xenografts. Neuro-oncology 2009.

[13] Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, Mischel PS, Nelson SF. Gene expression profiling of gliomas strongly predicts survival. Cancer research 2004;64(18):6503-10.

[14] Liang Y, Diehn M, Watson N, Bollen AW, Aldape KD, Nicholas MK, Lamborn KR, Berger MS, Botstein D, Brown PO, Israel MA. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. Proceedings of the National Academy of Sciences of the United States of America 2005;102(16):5814-9.

[15] Wiens DJ, Mann TK, Fedderson DE, Rathmell WK, Franck BH. Early heart development in the chick embryo: effects of isotretinoin on cell proliferation, alpha-actin synthesis, and development of contractions. Differentiation; research in biological diversity 1992;51(2):105-12.