Identifying biomarkers

Firstly, can you spell out the overall aims and objectives of your research programme in developing a novel computational framework for individualised clinical decision?

We would like to develop an innovative computational framework for the identification of new disease biomarkers from the whole human genome. Using the expression profiles of the identified biomarkers, diagnostic and prognostic models will be constructed to predict clinical outcome in individual patients. The constructed models could be used to advise physicians about treatment options for individual patients: for instance, whether or not a patient should receive adjuvant chemotherapy.

Why is the study and use of biomarkers important with regard to lung cancer recurrence?

Surgical resection is the major treatment option for stage I non-small cell lung cancer (NSCLC) patients. However, about 35-50 per cent of stage I, NSCLC patients will develop and die from recurrence within five years of surgery. The emerging use of biomarkers may enable clinicians to make treatment decisions based on the specific genetic makeup of individual patients.

Increasing evidence has shown that genetic markers could be used to identify more aggressive tumours.

Can you outline the advances made to date in refining the prognosis of NSCLC through the analysis of gene expression patterns (biomarkers) in tumours?

There have been significant advances in refining the prognosis of NSCLC by analysing gene expression patterns in tumours. Gene expression-based diagnosis of lung adenocarcinomas has already been incorporated in clinical settings to treat this deadly disease. We have recently identified and validated a 35-gene set from publicly available gene expression profiles of 434 non-small cell lung cancer patients. This signature is independent of other prognostic factors for non-small cell lung cancer, including age, sex, tumour differentiation, tumour grade, and tumour stage. This signature also had better prognostic performance than other lung cancer signatures on the studied cohorts.

Protein expression of two signature genes, TAL2 and ILF3, was confirmed in lung adenocarcinoma tumours. How significant is it that these two biomarkers showed elevated gene and protein expression levels in lung cancer development and progression; and in what way(s) might this outcome shape future research?

This finding is important. It implies that TAL2 and ILF3 might be functionally involved in lung cancer development and progression. We will perform biological experiments to validate this. In addition, we will test the protein expression level of these two genes in patient blood samples to see if they could potentially be used as markers for diagnosis and prognosis of lung cancer.

How might this project act as a platform in supporting or assisting future studies regarding lung cancer recurrence?

The finding of the project (the identified biomarkers) provides potential drug targets for treating lung cancer. The identified gene signatures could be used in clinics to select patients at high-risk for recurrence for adjuvant chemotherapy. The methodology could be used for biomarker identification in general.

Can you identify any particular difficulties or obstacles your team has faced during the project?

In the beginning, we did not have sufficient funding to perform wet-lab experiments to validate the identified biomarkers using patient samples. Now with the funding support to do this, the current challenge is to perform experimental assays up to the standards (and precision) in clinical tests.

Would you like to delineate any particular efforts or contributions of members of your team?

Ying-Wooi Wan is a PhD candidate in Computer Science. She performed the computational analysis and developed software in this project. Joseph Putila is a PhD candidate in Public Health. He analysed large population SEER data to incorporate traditional prognostic factors with gene signature. Rebecca Raese is a lab technician who performed RT-PCR analysis to validate the biomarkers. Our clinical collaborators Dr Afshin Dowlati (Case Western), Dr Scot Remick and Dr Barbara Ducatman at WVU provided substantial medical support. Dr David Beer from the University of Michigan provides essential expertise in translational research for this project. Dr Vince Castranova, Dr Yong Qian and Dr Val Vallyathan at the National Institute of Occupational Safety and Health provide laboratory space and facilities for the experiments. Dr Thomas Ried from the National Cancer Institute has worked with us in colorectal cancer and breast cancer using the proposed computational framework.
LUNG CANCER IS currently responsible for more deaths than breast, colon and prostate cancer combined. Alarmingly, when compared to other cancers, the amount of research and funding that lung cancer receives is disproportionate to its survival rates. As the leading cause of cancer deaths for both men and women, research into this deadly disease is proving vital.

Non-small cell lung cancer (NSCLC) accounts for about 80 per cent of incidences of the disease, growing and spreading more slowly than small cell lung cancer. NSCLC however has three different forms: adenocarcinomas, which are often found in an outer area of the lung; squamous cell carcinomas, usually found in the centre of the lung by a bronchus; and large cell carcinomas, which can occur in any part of the lung.

Alarmingly, 35-50 per cent of patients with stage I NSCLC relapse within five years. It therefore presents clinicians with the unanswered challenge of accurately estimating the risk of recurrence in individual patients. If a solution were found it would help to provide personalised therapy for individual patients, with a large subgroup of patients potentially benefiting from adjuvant chemotherapy.

Biomarkers - namely, biochemical features that enable the progression of a disease or the effects of its treatment to be measured - are emerging as a new method to help clinicians assess treatment decisions based on the specific genetic makeup of individual patients rather than more general results. NSCLC is an area that desperately needs this vital technology, however, personalised therapy requires innovative genome-scale studies in order to identify expression patterns in disease progression. Furthermore, each individual feature selection algorithm (a step by step computational procedure for solving a problem) has different strengths, so hybrid models that combine multiple algorithms, have become crucial to the identification of clinically relevant biomarkers.

Dr Nancy Lan Guo, from the Mary Babb Randolph Cancer Center of West Virginia University, has recognised the importance of revealing disease-mediated biomarker interactions, including feedback circuits, to develop more suitable and effective therapies to both predict metastasis and relapse in non-small cell lung cancer and colorectal cancer.

Lan Guo sees the need to create an innovative computational bioinformatics framework to facilitate this vision, by combining genomics and proteomics, as well as clinical approaches. This research will provide an accurate gene expression-based predictor of lung cancer recurrence, and will facilitate a much broader reaching and more informed decision making process. Not only could this genetic predictor allow patients to be categorised into either high or low risk brackets, but it would also offer a more improved cancer management structure for clinicians, therefore impacting current survival rates.

THE CHALLENGE

The human genome contains about 20,000 to 25,000 genes. Using current DNA microarray technology, expression levels of 54,000 probes are quantified on a single sample. This poses what Lan Guo refers to as a ‘curse of dimensionality’ of the data. By using hybrid models that combine multiple algorithms, new gene selection schemes to manage this problem can be developed, allowing clinically relevant biomarkers to be readily identified.

Different feature selection algorithms have different assumptions and strengths, which means that not all are able to identify a predictive gene set that could be used for clinical applications. For example, some algorithms can be used to select the most predictive biomarkers from a list of 50-200 candidate genes, but they are unable to process more gene variables. Lan Guo’s goal is to develop a feature selection system by combining multiple algorithms to generate the highest prediction accuracy with the minimum number of biomarkers. These will be determined for several different cancer types, with the identified biomarkers being appropriately validated by extensive public data sets. With this information Lan Guo and her team then wish to create an innovative framework for modelling the interaction of biomarkers for clinical classification. This network will serve to clarify these molecular interactions among the biomarker proteins, providing vital insight into the progression of the disease.

FRAMING GENES

Lan Guo’s computational framework will be based on the Dempster-Shafer theory. This hypothesis is founded on belief functions that allow degrees of belief to be based for one question on probabilities for a related question. It is worth noting that these degrees of belief do not necessarily have to be based on the mathematical properties of probabilities, but rather on how closely the two questions are related. Algorithms will therefore be used to optimise the performance of this network. The theory will be implemented in the belief networks to confront a variety of real life
clinical applications. Lan Guo and her team see the importance of then comparing and testing this network with others by using the same data sets, and evaluating the best molecular classifiers.

By exploring and developing different algorithms in different stages of gene selection, Lan Guo and her research team have been able to identify new prognostic gene signatures for both breast and lung cancer. They have also identified and validated 35-gene signature from gene expression profiles of 434 non-small cell lung cancer patients using gene DNA microarrays. These 35 genes were then confirmed with real-time RT-PCR analysis of 90 independent lung cancer tumour specimens. Lan Guo is validating the signature using RT-PCR low density arrays, since they provide a more accurate platform than microarray and are more appropriate for clinical application. She foresees the impact of such a development to be significant: “Once this signature is confirmed in RT-PCR analysis of independent patient samples, it would be a step forward to clinical utility and its clinical utility. The use of computer-based technologies is becoming increasingly important in biomarker research and drug discovery methods, with advances in high-throughput technologies facilitating improved screening of the human genome. Lan Guo also pre-empts the potential that web-based infrastructures hold for clinical decision making, expressing the possibility of downloading the software to physicians PDAs in the near future.

Lan Guo’s current research into gene signature is independent of factors for non-small cell lung cancer, including age, sex, tumour differentiation, tumour grade, and tumour stage. In the future, Lan Guo and her team wish to incorporate these traditional prognostic factors and gene expression signatures for patient stratification. This research is currently funded by a 1 million US dollar recovery fund, showing that there is huge potential for both biomarker research and its clinical utility. The use of computer-based technologies is becoming increasingly important in biomarker research and drug discovery methods, with advances in high-throughput technologies facilitating improved screening of the human genome. Lan Guo also pre-empts the potential that web-based infrastructures hold for clinical decision making, expressing the possibility of downloading the software to physicians PDAs in the near future.

Lan Guo’s approach truly emphasises the growing nature of multidisciplinary research, as her project involves collaborations from bioinformaticians, clinicians, and biomedical researchers for algorithm development. By developing an innovative computational framework, Lan Guo is blending these disciplines in a bid to identify new disease biomarkers from whole human genome. In proving this ‘genetic predictor’, Lan Guo’s research holds the potential not only to better manage lung cancer on an individual level, but to limit and prevent its devastating effects.

FIGURE 2. (A and B) PROTEIN EXPRESSION IN LUNG ADENOCARCINOMA TUMOUR TISSUES. Both lung cancer tumour tissues (T) and the adjacent normal tissues (N) were lysed. The tissue lysates were resolved by SDS-PAGE, followed by transferring to PVDF by Western transfer. The transferred proteins on PVDF were probed with the specific antibodies, TAL2 (A) and ILF3 (B), respectively. To confirm the specificity of the identified proteins, each protein was sequentially probed with two different antibodies (1 and 2) targeting at the different epitopes. The transferred PVDF was probed with the first round antibodies (1), and then was stripped with the stripping buffer, followed by probing with the second round antibodies (2).

we would consider this our greatest success to date in this research area” she states. Lan Guo’s research has confirmed the protein expression of two signature genes, TAL2 and ILF3, in lung adenocarcinoma tumours. These two biomarkers demonstrated elevated gene and protein expression levels in both the development of lung cancer and in its progression.

Lan Guo’s current research into gene signature is independent of factors for non-small cell lung cancer, including age, sex, tumour differentiation, tumour grade, and tumour stage. In the future, Lan Guo and her team wish to incorporate these traditional prognostic factors and gene expression signatures for patient stratification. This research is currently funded by a 1 million US dollar recovery fund, showing that there is huge potential for both biomarker research and its clinical utility. The use of computer-based technologies is becoming increasingly important in biomarker research and drug discovery methods, with advances in high-throughput technologies facilitating improved screening of the human genome. Lan Guo also pre-empts the potential that web-based infrastructures hold for clinical decision making, expressing the possibility of downloading the software to physicians PDAs in the near future.

Lan Guo’s approach truly emphasises the growing nature of multidisciplinary research, as her project involves collaborations from bioinformaticians, clinicians, and biomedical researchers for algorithm development. By developing an innovative computational framework, Lan Guo is blending these disciplines in a bid to identify new disease biomarkers from whole human genome. In proving this ‘genetic predictor’, Lan Guo’s research holds the potential not only to better manage lung cancer on an individual level, but to limit and prevent its devastating effects.

FIGURE 2. (A and B) PROTEIN EXPRESSION IN LUNG ADENOCARCINOMA TUMOUR TISSUES. Both lung cancer tumour tissues (T) and the adjacent normal tissues (N) were lysed. The tissue lysates were resolved by SDS-PAGE, followed by transferring to PVDF by Western transfer. The transferred proteins on PVDF were probed with the specific antibodies, TAL2 (A) and ILF3 (B), respectively. To confirm the specificity of the identified proteins, each protein was sequentially probed with two different antibodies (1 and 2) targeting at the different epitopes. The transferred PVDF was probed with the first round antibodies (1), and then was stripped with the stripping buffer, followed by probing with the second round antibodies (2).