'CHIP'ping a way through the Cancer Maze

Over the past several decades the ability to distinguish and reproducibly score differences between normal and malignant cells has sharpened progressively. Much of this has resulted from the exploitation and translation of knowledge gained from research benches to the bedside. In the course of this journey there has been a gradual shift from discerning gross pathological features, histological characters, specific cell surface markers to the use of non-random genetic markers in identifying and classifying malignancies.

The most visible impact of translating 'bench' technologies to the 'bedside' understanding of cancer has been in the field of lymphoid malignancies. One consequence of this is the constant and dynamic reappraisal of classification schemes. The present recommendations of classification of lymphoid malignancies, for example, while seemingly too exhaustive to the practising oncologist, nonetheless highlight the complexity and the interrelationships in the biology of what some years ago were just a group of four broad cancers—Hodgkin's disease, lymphosarcoma, reticulum cell sarcoma and giant follicle lymphoma.

However, it is also clear from the unpredictability of clinical outcomes in the presently defined subgroup of diffuse large B cell lymphomas that even the present classification is underpowered and cannot distinguish between some subgroups of histological, immunophenotypic and, within the limitations of applicable molecular variables, apparently identical lymphomas. These concerns, along with the recognition of the multiplicity of molecular underpinnings that empower the cancer cell, have provided the impetus for developing innovative and comprehensive methods of clinically relevant categorization of cancer.

The observable phenotypes of malignant lymphoid cells, be it morphology, expression of cell surface markers, proliferative capacity, apoptotic indices or interactions with the micro-environment, are eventually a summation of the concerted action of multiple genetic pathways involved in controlling growth, death, differentiation and cell–cell interactions. Therefore, portraying genome-wide expression of genes in lymphoid cells should ultimately provide molecular fingerprints of expression that comprehensively distinguish these phenotypes.

In what may seem more like science fiction, the newly developed tools are empowered with the capabilities to catalogue genome-wide and global transcriptional activity of cancer cells.

Not so long ago, molecular oncologists used Northern blot analysis to study the relative expression of single genes from tumour samples. Such studies gave way to utilizing a reverse Northern blot, in which 'probes' representing multiple genes, rather than the test RNA, were immobilized and the transcripts from a tumour were labelled 'in toto' and hybridized to the reverse blot, thus allowing the interrogation of the expression of multiple genes in a given tumour sample.

The genomic revolution allowed the expansion of the number of unique gene sequences available for such studies, and the parallel improvements in technology allowed the 'printing' of several thousands of these gene sequences on a single hybridizable matrix in dimensions that can be practically handled.

It is apparent that the bewildering quantity of raw data that such an approach can
generate would be useful only if it is coupled with decisive computational approaches that can identify distinctive transcriptosomal signatures between different cancer tissues.\(^7\)\(^-\)\(^10\)

Golub et al.\(^1\) initially provided the test of such a marriage—between computational bioinformatics and gene hybridization technologies—to conceive clinically relevant and practical information in distinguishing leukaemias.

Among leukaemias, two biologically distinct categories—acute myeloid (AML) and acute lymphoblastic leukaemias (ALL)—arise independently from myeloid or lymphoid precursors. Consequently, these can frequently be differentiated using contemporary tools of immunophenotyping and cytogenetics. Nonetheless, the possibility of errors in diagnosis do exist and since distinct treatment strategies are required for AML and ALL, these errors can be detrimental. The unique features that allow a differential diagnosis of myeloid and lymphoid leukaemia must be replicated in the overall profiles of gene expression in these two leukaemic subgroups.

Golub et al. tested this hypothesis by analysing the levels of transcription of 6800 genes in ALL and AML samples obtained from patients. The group also developed a novel computational approach to handle the raw data obtained from the expression study. The approach was to identify gene expression patterns in samples previously validated as AML or ALL by established tools. Over a thousand genes, including those that could be predicted from the existing information of ALL and AML, qualified as candidates whose expression patterns correlated with the leukaemia type. These include genes such as CD33 that encode for cell surface markers presently useful in defining myeloid versus lymphoid cell types.

The investigators then created a subset of the most informative genes to form a class predictor set of 50 genes that could be closely correlated with distinction of AML and ALL. This set of class predictor genes were cross-validated using known samples. Comparisons between the level of expression of each gene projected in the predictor set and the observed level of expression in the test leukaemia were used to make the diagnosis of AML or ALL for the test sample. In 38 leukaemias thus tested, 36 could be correctly assigned. Further validation of their computational algorithms were made by analysing an additional set of 34 leukaemia samples obtained from diverse laboratories that used different sample preparation procedures. The class predictor set of 50 genes correctly distinguished 29 of the 34 samples. The next phase of the test, which the investigators label as class discovery, was to determine if the expression of the set of 50 genes could be used to automatically discover the two classes of leukaemia—AML and ALL—without any prior knowledge of the distinction. Indeed, the success of this test is what makes the approach remarkable. The ability of combined gene expression and the algorithms to cluster tumour types allowed the investigators to successfully label diverse samples in a real clinical setting. Notably this was achievable independent of multiple separate tests carried out in several laboratories, as is routinely done to identify AML and ALL.

The impetus to the development of these technologies as discovery tools for classifying cancer based upon molecular fingerprints was directly derived from the Cancer Genome Anatomy Project (CGAP) of the National Cancer Institute (NCI) that created a tumour gene index to identify and sequence cDNAs expressed in different tumours. Dr Louis Staudt, at the NCI, for instance, derived from the CGAP database over 15 000 genes that were expressed in B lymphocytes. These, along with an additional 300 genes, were used to construct a lymphochip.\(^12\)

The diffuse large B cell group of lymphomas appear to be heterogeneous, based upon clinical outcome to uniform therapy. However, any subclassification of these lymphomas is not evident by traditional histopathological methods or immunophenotyping. When these tumours were classified using cluster-based analysis of expression data obtained using the lymphochip, two distinct subgroups were evident. In one the expression pattern was similar to that seen in germinal centre B cells and in the other the expression pattern was similar to that obtained in B cells induced to activate in vitro. More important, these two groups also differed in clinical outcome;
76% of patients with the germinal centre-like B cell lymphoma survived for 5 years compared to 16% for the other group.

There are numerous other ongoing studies that are being carried out to answer several burning questions. Are there common aberrant pathways in leukaemias that are otherwise distinct, for example, precursor B cell ALL with different chromosomal translocations? Can target genes be identified that can predict relapse? Are there common pathways that determine poor clinical outcome in otherwise ‘good’ leukaemias such as the TEL-AML and leukaemias that frequently respond poorly, example BCR-ABL? Comparison of cytogenetically distinct subtypes of lymphoid and myeloid leukaemias with normal counterparts should also provide important information on the totality of corrupted pathways.

Yeoh et al.\textsuperscript{13} recently presented the results of profiling paediatric ALL at the American Society of Hematology meeting. The cluster analysis impressively segregated known subgroups of ALL based upon their expression profiles, the prediction accuracies of the computational analysis reached 100%, and accurate predictions of T cell ALL and precursor B cell ALL could be made. Additionally, the profiles neatly integrated ALL with unique chromosomal abnormalities. Thus, the expression pattern could distinguish precursor B cell ALL with E2A-PBX, TEL-AML, BCR-ABL or MLL rearrangements. Acute lymphoblastic leukaemia with hyperdiploidy was also recognized as an entity with a unique expression pattern.

Similar studies in other haematopoietic malignancies also validated molecular signature approaches as powerful classification tools. Schoch et al.\textsuperscript{14} demonstrated distinct gene expression patterns in subgroups of AML. These patterns of expression correlate with abnormalities at the genomic level. Interestingly, this study highlighted the expression of 2–7 genes that would sufficiently define four subgroups of AML.

It is evident that in addition to the deregulated oncogenes, the cell type significantly influences the expression patterns. Thus, distinct global expression patterns were indicated when TEL-AML and AML-ETO leukaemias were compared. As would be predicted, a major function of upregulated genes could be freely associated with granulocyte function, most notably the granulocyte enzymes in AML-ETO leukaemias, while B lymphocyte-associated proteins such as CD10, CD19 and CD79a figured in TEL-AML leukaemias.

The profile analysis of paediatric T cell ALL presented by Ferrando et al.\textsuperscript{15} using the Affymetrix HU6800 array, clearly distinguished distinct classes of T cell ALL, in the context of thymocyte development. T cell ALL associated with HOX-11 demonstrated an expression pattern distinct from the LYL-1 T cell ALL and TAL-1 T cell ALL. While gene expression patterns in HOX-11 associated T cell ALL indicated a cortical thymocyte progenitor cell, TAL-1 T cell ALL was clearly associated with the late cortical thymocyte stage of development. Additionally, a distinct subclass that was not associated with LYL-1, TAL-1 or HOX-11 could be distinguished based upon profile analysis indicating the ‘class discover’ potential of profile approaches.

A finding of major interest in the study by Yeoh et al.\textsuperscript{13} was the identification of a predictor set of genes with the property to risk-stratify patients by identifying patients at risk for relapse and those likely to develop therapy-related secondary malignancies.

Overall, these studies demonstrate the immense power of genomic profiling in defining clinically relevant signatures of malignant cells. It is clear that no single gene but a combination of genes will provide the best predictive value and that more often than not, a direct relationship between the expression profile and a specific chromosomal translocation can be established. This single platform should thus allow assignment of malignancies into known classes and generate novel classes that are clinically relevant.

While these studies demonstrate the power and feasibility of developing clinically relevant molecular portraits of cancer, several hurdles remain including the cost of microarray facilities and the availability of experienced computational and bioinformatics scientists. However, it is likely that in parallel with these hurdles being overcome, other complementary technologies will also develop, including
versions of microarray assessments that use fluidic technologies. These should also provide the potential for real-time profiling of cancer signatures. The widespread use of microarray profiling in routine clinical practice will only result from further innovations that enable the generation of a cost-effective, automated hand-held device that provides the clinician with a molecular portrait of the individual patient’s malignancy and help the oncologist determine the appropriate individualized therapeutic intervention. How the ‘CHIP’ will help improve ‘CHOP’ will certainly be an interesting story worth following.

REFERENCES


