Clinical implementation of the intrinsic subtypes of breast cancer

Our group has reported several intrinsic gene sets important for identifying subtypes of breast cancer with clinical significance.1–3 In these studies we have explored and published methods for sample classification across different genomic platforms and tissue qualities. In 2009, we suggested the use of a standardised gene set (PAM50) for subtype classification to improve the classification concordance reported by investigators.3 However, a standardised gene set does not completely resolve discrepancies between researchers since the genes might be quantitatively measured using different platforms and normalisation methods. Weigelt and co-workers4 applied three different intrinsic gene sets to four data sets using one prediction method and showed a range of agreement. Because of the level of discordance that was reported, they concluded that identification of the intrinsic subtypes is not ready for clinical implementation. We disagree with this interpretation.

Careful examination of Weigelt and co-workers’ analyses4 revealed bioinformatics-based technical limitations that reduced the accuracy in subtype predictions and concordance of these three predictors. These limitations are highlighted in the accompanying letters, but we emphasise here the importance of dataset to dataset normalisation. The webappendix shows the relationship between the four datasets when they are not normalised (as done by Weigelt and co-workers4) and when row centring and column standardisation is done (as advocated by ourselves and others1–3,5,6). Additionally, differences in the composition of datasets (ie, proportion of oestrogen receptor-positive [ER+] tumours) can affect sample classification.2 Nonetheless, all three gene sets were significant predictors of outcomes in univariate and multivariate testing, which suggests that this is a robust classification method.

Many clinical assays begin in the research setting and are refined over time until ready for clinical use. Clinical concordance testing rarely has perfect agreement even under the best of circumstances, such as measuring a single analyte with a locked-down protocol across CLIA (Clinical Laboratory Improvement Amendments) laboratories. There is little value, and potential harm, to draw conclusions about the robustness and utility of a test based on research data from independent laboratories not intended for concordance testing, as Weigelt and colleagues4 and the accompanying commentary5 interpreted their findings.

The interpretation of Weigelt and co-workers4 is also based on the hypothesis that training sets with different tumours and genes should result in high agreement in subtype classification. In fact, these three training sets were not specifically designed to be concordant at the individual-sample level. The reason for this is that these three classifiers reflect the logical evolution over time of a classification method based on the most up-to-date data and technologies available. Over the last decade, we have learned a great deal about microarray experimental design, objective statistical selection procedures, and the microarray technology itself has dramatically improved. For these reasons, we believe that the most accurate and trustworthy assay is the PAM50 assay.3

In view of the confusion over what a single sample predictor (SSP) is, we define a SSP here as any predictor where the algorithm and any parameter values are exclusively determined from a training set, and test cases are assessed independently. This requires that normalisation of a test case (be it R/G ratio or housekeeper normalised for example) is not dependent on measurements from other test cases. In theory, any centroid predictor could be used as a SSP; however, in reality, this is only practical when a single platform is used for both training and testing predictions (as with the qRT-PCR PAM50, Oncotype DX [Genomic Health, USA], and Agenda Mammaprint assays [Agenda, Netherlands]). This is because different protocols and assays for the same gene will exhibit measurement bias. Measurement bias could be reduced with controls, but controls are often unavailable across public datasets. In these cases, normalisation estimates (ie, gene centring) must be determined from observations of the test cases and this precludes true SSP.

We stand by the performance of the old and current versions of the intrinsic subtyping assays, and feel that the prognostic and predictive abilities of the current PAM50 assay will be ready for clinical implementation after further validation on additional clinical cohorts. The current PAM50 assay is the culmination of more than 10 years of research, and was optimised on the basis of new data and many lessons learned from previous publications. For example, the PAM50 assay is the only
assay that was based on a pure prognostic population of patients with breast cancer, and is the only one to be trained on data from two gene-expression platforms (microarray and qRT-PCR); this assay has shown potential for treatment prediction. For these reasons, we intend to pursue further the clinical validation of the PAM50 assay within a single platform through retrospective and prospective clinical-trial analyses.

Charles M Perou*, Joel S Parker, Aleix Prat, Matthew J Ellis, Philip S Bernard
Lineberger Comprehensive Cancer Center (CMP, JSP, AP), Department of Genetics (CMP, JSP, AP), Department of Pathology & Laboratory Medicine (CMP), University of North Carolina, Chapel Hill, NC 2759, USA; Division of Oncology, Department of Medicine, Washington University School of Medicine, St Louis, MO, USA (MJE); Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA (PSB)
cperou@med.unc.edu

The importance of gene-centring microarray data

In the April, 2010, issue of The Lancet Oncology, Weigelt and colleagues1 investigate the association between three different methods of predicting molecular subtypes of breast cancer, all referred to as single sample predictors (SSPs). Their conclusions, however, are flawed by the use of uncentred microarray data. The original methods are all based on their correlation to expression centroids generated from three different gene lists. Weigelt and colleagues apply the methods to four different gene-expression datasets without performing the essential gene centring procedures before classification. The consistency between these subtype assignments was measured by Cohen’s kappa scores. Several important conclusions were made on the basis of these erroneous classifications: different SSPs produce inconsistent results, none of the microdissected specimens with a tumour-cell content greater than 90% were assigned to the normal breast-like group, and the human epidermal growth factor receptor 2 (HER2; also ERBB2) group, as defined by microarray analysis, does not equate with the clinical subgroup of HER2-positive breast cancer.

The effect and importance of centring is shown in the webappendix for the Sørlie centroids; similar results apply to the Hu and PAM centroids (data not shown). For the two-channel NKI-295 data (webappendix C), the centred and uncentred centroid correlations correspond reasonably well since the comparison with a common reference suggests partial centring. For the remaining three one-channel-based datasets (webappendix A,B,D), most of the variation is caused by differences in the general expression level of different genes, and not as much by differences between samples. Hence, for uncentred data, correlation of the expression level to the centroids biases the results by raising correlations to the luminal-B centroid and lowering correlations to the normal-like centroid, which explains why many samples are classified as luminal B and few as normal-like. Also, the correlation values vary over a smaller range in the uncentred data, because the sample differences only constitute a small portion of the variance.

The subtype centroids from the original classification,2–4 are based on median-centred two-channel microarray data. For a sample to be correctly assigned to a subtype, it must be centred against an appropriately large and heterogeneous sample set. This is fundamental when applying the classifier to samples handled by different expression platforms rather than the original dataset5 and hence disqualifies the Sørlie approach from being used as a SSP in the sense that Weigelt and colleagues1 did. Applying this method to uncentred data cannot be expected to give meaningful results. The Hu6 and Parker7 methodology is not a SSP in the sense that Weigelt and colleagues1 did. Applying this method to uncentred data cannot be expected to give meaningful results. The Hu6 and Parker7 method is not a SSP in the sense that Weigelt and colleagues1 did. Applying this method to uncentred data cannot be expected to give meaningful results. The Hu6 and Parker7 assay is based on a pure prognostic population of patients with breast cancer, and is the only one to be trained on data from two gene-expression platforms (microarray and qRT-PCR); this assay has shown potential for treatment prediction. For these reasons, we intend to pursue further the clinical validation of the PAM50 assay within a single platform through retrospective and prospective clinical-trial analyses.

Charles M Perou*, Joel S Parker, Aleix Prat, Matthew J Ellis, Philip S Bernard
Lineberger Comprehensive Cancer Center (CMP, JSP, AP), Department of Genetics (CMP, JSP, AP), Department of Pathology & Laboratory Medicine (CMP), University of North Carolina, Chapel Hill, NC 2759, USA; Division of Oncology, Department of Medicine, Washington University School of Medicine, St Louis, MO, USA (MJE); Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA (PSB)
cperou@med.unc.edu

CMP, MJE, and PSB are co-founders of University Genomics and are major stock holders of University Genomics and Biodclassifer LLC. CMP, MJE, PSB, and JSP have filed a patent for the PAM50 assay from the University of North Carolina and University of Utah. AP declared no conflict of interest.