Methodological quality assessment of studies investigating the PAM50 test

Study feature	Qualities sought	Bernard *et al.* (2011)[119] (abstract: analytical validity) [Additional data from Martin *et al.*[120] (abstract: clinical utility)]	Cheang *et al.* (2011)[117]	Chia *et al.* (2011)[121]	Ebbert *et al.* (2011)[118] (abstract)	Nielsen *et al.* (2010)[116]	Parker *et al.* (2009)[115]
Sample of patients	Inclusion criteria defined	U	Y	Y	U	Y	Y
	Sample selection explained	U	Y	Y	U (breast samples – no further details provided)	Y	Y
	Adequate description of diagnostic criteria	U	Y	Y	U	Y	Y
	Clinical and demographic characteristics fully described	U	Y	Y	U	Y	Y
	Representative (random or consecutive sample)	U	Y (random – data and samples from a RCT)	Y (random – data and samples from a RCT)	U	U (cohort – cases with complete outcomes and representative sample)	U (cohort – tissues collected under approved protocols)
	Assembled at a common (usually early) point in the course of their disease	U (all LN+)	Y (all LN+)	U (75% LN+)	U	U (65% LN+)	U (>85% LN–)
	Complete (all eligible patients were included)	U	Y	N	U	N	U

Study feature	Qualities sought	Bernard *et al.* (2011)[119] (abstract: analytical validity) [Additional data from Martin *et al.*[120] (abstract: clinical utility)]	Cheang *et al.* (2011)[117]	Chia *et al.* (2011)[121]	Ebbert *et al.* (2011)[118] (abstract)	Nielsen *et al.* (2010)[116]	Parker *et al.* (2009)[115]
Follow-up of patients	Sufficiently long	Y (8.7 years)	Y	Y (median 9.7 years)	NA	Y	U
Outcome	Objective	Y	Y	Y	NA	Y	Y
	Unbiased (e.g. assessment blinded to prognostic information)	U	Y	U	NA	U	U
	Fully defined	U	Y	N	NA	Y	N
	Appropriate	Y (DFS, OS)	Y (RFS, OS)	Y (RFS, OS)	NA (analytical)	Y (RFS, DSS)	Y (RFS)
	Known for all or a high proportion of patients	U	Y	Y	NA	Y	Y
Prognostic variable	Fully defined, including details of method of measurement if relevant	U	Y	Y	U	Y	Y
	Precisely measured	U	Y	Y	U	Y	Y
	Available for all or a high proportion of patients	U	Y	Y	U	Y	Y
	If relevant, cut-point(s) defined and justified	U	Y (detailed)	Y (detailed)	U	Y (detailed)	Y (provided reference)
Analysis	Continuous predictor variable analysed appropriately	U	Y	Y	U	Y	Y
	Statistical adjustment for all important prognostic factors	U	Y (includes multivariate analysis)	Y (includes multivariate analysis)	U	Y (includes multivariate analysis)	Y (includes multivariate analysis)
Intervention subsequent to inclusion in cohort	Fully described	U	Y	Y	U	N	N
	Intervention standardised or randomised	U	Y	Y	U	N	N

DSS, disease-specific survival; N, no; NA, not applicable; U, unclear/not reported; Y, yes.

Summary of results: PAM50 test

Study	Outcomes/end points	Results	Authors Conclusions	Comments
Bernard *et al.*, 2011[119] (abstract) Additional data from Martin *et al.*[120] (abstract)	Analytical outcomes including accuracy and reproducibility	**Bernard *et al.*[119]** There was good agreement between RT-qPCR gene expression and IHC scoring for the clinical markers (gene/protein) *ESR1*/ER, *PGR*/PR and *ERBB2*/HER2. The accuracy was significantly lower for *MKI67*/Ki-67, *EGFR*/EGFR and *KRT5*/CK5/6. Discrepancies between the Herceptest and CISH for a test score of 2+ and 3+ samples showed that RT-qPCR agreed better with the Herceptest (AUC 0.95 vs. 0.93) **Martin *et al.*[120]** Concerning predictive factors, exploratory analyses showed that fluorouracil, epirubicin, cyclophosphamide and paclitaxel (FEC-P) was better than FEC in the low PR group (HR: 0.68, $p=0.033$) and not in the high PR group (HR: 0.83, $p=0.245$); interaction test $p=0.358$. Similarly, FEC-P was better in the low *ERBB2* group (HR: 0.67, $p=0.005$) and not in the high *ERBB2* group (HR: 0.92, $p=0.707$); interaction test $p=0.256$. In addition, superiority of FEC-P was observed for the low proliferation signature group (HR: 0.58, $p=0.014$) in contrast to the high proliferation signature group (HR: 0.93, $p=0.633$); interaction test $p=0.069$. The FEC-P group showed improved outcomes in all genomic intrinsic subtypes, although no subtype alone reached statistical significance	Calling cut-points based on RT-qPCR expression across subtypes is reproducible across data sets and has good agreement with expression by IHC for clinically used biomarkers. In addition, the PAM50 proliferation signature could be predictive of benefit for adding weekly paclitaxel to the adjuvant chemotherapy FEC regimen. These results need further validation in an independent study	

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors Conclusions</th><th>Comments</th></tr>
<tr><td>Cheang et al. (2011)[117] (additional data from unpublished manuscript)</td><td>Responsiveness of intrinsic subtypes to adjuvant anthracyclines vs. non-anthracyclines</td><td>Intrinsic subtyping using the PAM50 assay
<table><tr><th>Subtype</th><th>Tumour samples analysed (n=476), n (%)</th></tr><tr><td>Luminal A</td><td>146 (30.7)</td></tr><tr><td>Luminal B</td><td>110 (23.1)</td></tr><tr><td>HER2 enriched</td><td>105 (22.1)</td></tr><tr><td>Basal-like</td><td>94 (19.7)</td></tr><tr><td>Normal-like</td><td>21 (4.4)</td></tr></table>Association of intrinsic subtypes with survival
Intrinsic subtypes were significantly associated with RFS ($p=0.0005$) and OS ($p<0.0001$) on the combined cohort. The HER2-enriched subtype demonstrated the greatest benefit from FEC vs. CMF, with an absolute difference of more than 20% in both 5-year RFS and OS, whereas there was a <2% difference for the non-HER2-enriched tumours (interaction $p=0.03$ for RFS and $p=0.02$ for OS). Within tumours defined clinically as HER2+ by IHC or fluorescence in situ hybridisation, 79% (72/91) were classified as the HER2-enriched subtype by genomics and these tumours were also significantly associated with better response to CEF vs. CMF (62% vs. 22%, $p=0.0006$). In contrast, basal-like tumours ($n=94$) did not benefit from the substitution of methotrexate for epirubicin with a HR of 1.1 for RFS and 1.3 for OS in favour of methotrexate, but the test for interaction was not significant</td><td>The HER2-enriched assignment strongly predicted anthracycline sensitivity. The chemotherapy-sensitive basal-like tumours showed no benefit for CEF, suggesting that non-anthracycline regimens should be further investigated in this subtype</td><td>Additional data available but not extracted</td></tr>
</table>

Study	Outcomes/end points	Results	Authors Conclusions	Comments
Chia *et al.* (2011)[121]	Intrinsic subtyping, disease-free survival, OS, risk of relapse modelling	**Intrinsic subtyping using the PAM50 assay** **Subtype** / **Tumour samples analysed ($n=398$), n (%)[a]** Luminal A: 135 (34) Luminal B: 76 (19) HER2 enriched: NR Basal-like: NR Normal-like: NR NR, not reported. a All numbers calculated based on reported percentages. **Intrinsic subtyping comparing the PAM50 assay with IHC** The concordance for intrinsic subtypes among 348 patients who could be classified by both the IHC and PAM50 classifiers was 70.8%, 86.8%, 80.2% and 93.4% for luminal A, luminal B, HER2 enriched (and basal-like breast cancers, respectively, with an overall kappa of 0.57 (95% CI 0.51 to 0.64) **Association of intrinsic subtypes with survival** Intrinsic subtypes as classified by the PAM50 assay were prognostic for both DFS ($p=0.0003$) and OS ($p=0.0002$), with the HER2-enriched subtype having the lowest and the luminal A subtype the highest 5-year survival values (DFS: 52.8% vs. 84.2%; OS: 68.1% vs. 95.7%). The prognostic value remained significant for both DFS ($p=0.02$) and OS ($p=0.02$) in multivariate analysis. Classification by the IHC panel was not statistically significant **Prediction of tamoxifen benefit** Luminal subtype by PAM50 was predictive of tamoxifen benefit (DFS: HR 0.52, 95% CI 0.32 to 0.86 vs. HR 0.80, 95% CI 0.50 to 1.29 for non-luminal subtype), although the interaction was not significant ($p=0.24$). Neither subtyping by central IHC nor by local ER status was predictive	Intrinsic subtype classification with the PAM50 assay was superior to IHC profiling for both prognosis and prediction of benefit from adjuvant tamoxifen	Additional data available but not extracted

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors Conclusions</th><th>Comments</th></tr>
<tr><td>Ebbert et al., (2011)[118] (abstract)</td><td>Analytical outcomes including accuracy and reproducibility</td><td>Within-platform cross-validation of the clinical subtype predictor showed 91.6% concordance. There was 100% reproducibility in subtype predictions across 46 runs testing different subtypes. Subtype predictions across platforms showed 88.1% concordance. Dilution experiments, introducing ‘normal’ breast tissue RNA into breast cancer RNA, showed a systematic switch towards the ‘normal’ signature, with luminal A and luminal B subtypes being most susceptible</td><td>The PAM50 Breast Cancer Intrinsic Classifier is highly reproducible within and across platforms. The clinical test has utility in the management of ER+ and ER– invasive breast cancers of all stages. It provides a necessary tool for identifying differences in tumour biology that are important for guiding patient care</td><td></td></tr>
<tr><td>Nielsen et al., (2010)[116]</td><td>Numbers assigned to each intrinsic subtype, risk of relapse score
Comparators: clinical, IHC (ER, PR, HER2, Ki-67)
Adjuvant! Online used to generate breast cancer RFS and disease-specific survival (DSS) estimates for each patient</td><td>Prognosis – intrinsic subtypes and risk of relapse by PAM50 and comparators
Intrinsic subtyping of ER+, tamoxifen-treated breast cancer using the PAM50 assay
<table><tr><th>Subtype</th><th>ER+, tamoxifen-treated systemic therapy (n=786), n (%)</th></tr><tr><td>Luminal A</td><td>372 (47.3)</td></tr><tr><td>Luminal B</td><td>329 (41.9)</td></tr><tr><td>HER2 enriched</td><td>64 (8.1)</td></tr><tr><td>Basal-like</td><td>5 (0.6)</td></tr><tr><td>Normal-like</td><td>16 (2.0)</td></tr></table>Kaplan–Meier survival analysis of intrinsic subtypes and ROR-S, as determined by PAM50
The included patients were considered to be at high risk with overall 10-year RFS of 62% and DSS of 72%. Those assigned to luminal A had significantly a better outcome (10-year RFS 74%; DSS 83%) than those assigned to luminal B, HER2-enriched and basal-like tumours
In Cox models incorporating standard prognostic variables, HRs for BCSS over the first 5 years of follow-up, relative to the most common luminal subtype, were 1.99 (95% CI 1.09 to 3.64) for the luminal B subtype, 3.65 (95% CI 1.64 to 8.16) for the HER2-enriched subtype and 17.71 (95% CI 1.71 to 183.33) for the basal-like subtype (p=0.0018)</td><td>C-index, Kaplan–Meier analysis and Cox model analyses show that IHC approaches do work and provide significant prognostic information; however, PAM50 is superior in terms of adding significant additional information and in its capacity to identify a particularly low-risk group. PAM50 can be applied to a large sample of FFPE breast cancers and gives more prognostic information than clinical factors and IHC</td><td>Authors highlight that the studied population was biased towards higher-risk breast cancers and thus underestimates the broader N0 population for whom adjuvant tamoxifen would represent adequate treatment
Additional data available but not extracted</td></tr>
</table>

ROR-S, risk of recurrence score based on subtype.

<table>
<tr><th>Study</th><th>Outcomes/end points</th><th>Results</th><th>Authors Conclusions</th><th>Comments</th></tr>
<tr><td>Parker et al. (2009)[115]</td><td>Distribution of instrinsic subtypes in comparison with clinical marker status
Risk of relapse models for prognosis in LN– breast cancer</td><td>Intrinsic subtyping using the PAM50 assay
<table><tr><th>Subtype</th><th>No adjuvant systemic therapy (n=761), n (%)[a]</th></tr><tr><td>Luminal A</td><td>269 (35.3)</td></tr><tr><td>Luminal B</td><td>168 (22.1)</td></tr><tr><td>HER2 enriched</td><td>120 (15.8)</td></tr><tr><td>Basal-like</td><td>128 (16.8)</td></tr><tr><td>Normal-like</td><td>76 (10.0)</td></tr></table>a All numbers calculated based on reported percentages.

Comparison of relapse prediction models
The intrinsic subtypes showed prognostic significance (for RFS) in untreated (no systemic therapy) patients and remained significant in multivariable analyses that incorporated clinical covariates (ER status, histological grade, tumour size and LN status)</td><td>The intrinsic subtype and risk predictors based on the PAM50 gene set added significant prognostic and predictive value to pathological staging, histological grade and standard clinical molecular markers</td><td>Neoadjuvant data available in paper but not extracted</td></tr>
</table>