supplementation with 0·5–5·0 mg folic acid lowered blood homocysteine concentrations by 25%. From our data, the shift in homocysteine dependency from folate to vitamin B12 with folic acid intervention becomes apparent only as the doses of folic-acid supplementation increase.

*Joe McPartlin, Helene McNulty, Eoin Quinlivan, Mary Ward, John Scott

*Department of Clinical Medicine, Trinity College, Dublin 8, Ireland; Department of Biochemistry, Trinity College, Dublin 2; and Northern Ireland Centre for Diet and Health (NICHE), School of Biomedical Sciences, University of Ulster, Coleraine, Londonderry, Northern Ireland, UK
(e-mail: jmcprtln@tcd.ie)

1 Homocysteine Lowering Trialists' Collaboration. Lowering blood homocysteine with folic acid based supplements: meta-analysis of randomised trials. *BMJ* 1998; **316:** 894–98.

Sir—E P Quinlivan and colleagues (Jan 19, p 227)[1] report a relation between total plasma homocysteine (tHcy) and vitamin B12 in folate-supplemented individuals. There are two B12 carrier proteins in serum. Haptocorrin binds most serum B12 but does not deliver the vitamin to metabolically active cells; this function is done by transcobalamin. Only 5–20% of serum B12 is bound to transcobalamin as holotranscobalamin. Current laboratory assays measure total serum B12 and are relatively poor indicators of the ability of serum to deliver the vitamin to tissues.[2]

We measured tHcy, folate, total serum B12, and holotranscobalamin concentrations in 111 elderly individuals (51 patients with dementia and 60 controls) with a mean age of 77 years (SD 9·5) recruited to a continuing study of vitamin B12 status and cognitive function. tHcy was measured by an automated high-performance liquid chromatography system, and folate and total serum B12 by an automated chemiluminescence analyser.

We used generalised linear models to investigate the relation between known tHcy determinants (age, sex, creatinine, smoking history, and serum folate) and B12 status, assessed by total serum B12 or holotranscobalamin concentrations. Diagnosis was included as an additional independent variable. Individuals receiving B vitamin supplements were excluded.

There was a significant and independent relation between tHcy and B12 status assessed by holotranscobalamin concentrations, but not by total serum B12 (table). Current laboratory assays lack sensitivity to measure biologically available B12. We suggest this might be an additional explanation as to why the effects of B12 on tHcy concentrations are frequently masked by folate status.

*Andrew McCaddon, Peter Hudson, Dick Ellis, Diane Hill, Alwyn Lloyd

*Gardden Road Surgery, Rhosllanerchrugog, Wrexham LL14 2EN, UK; University of Wales College of Medicine, Cardiff, UK; Wrexham Maelor Hospital, UK; University Hospital of Wales, Cardiff; Drew Scientific Ltd, Barrow-in-Furness, UK; and Wrexham Maelor Hospital, Wrexham
(e-mail: andrew@mccaddon.demon.co.uk)

1 Quinlivan EP, McPartlin J, McNulty H, et al. Importance of both folic acid and vitamin B12 in reduction of risk of vascular disease. *Lancet* 2002; **359:** 227–28.
2 Wickramasinghe SN, Fida S. Correlations between holo-transcobalamin II, holo-haptocorrin, and total B12 in serum samples from healthy subjects and patients. *J Clin Pathol* 1993; **46:** 537–39.

Variable	GLM including total serum B12		GLM including holotranscobalamin	
	Estimate (95% CI)	p	Estimate (95% CI)	p
Dementia	–0·86 (–1·48 to –0·23)	0·007	–0·57 (–1·2 to 0·02)	0·06
Sex	–0·10 (–0·82 to 0·62)	0·78	–0·20 (–0·93 to 0·52)	0·57
Smoker	–0·26 (–1·12 to 0·58)	0·54	–0·23 (–1·07 to 0·61)	0·59
Age	–0·01 (–0·09 to 0·08)	0·87	0·01 (–0·07 to 0·08)	0·89
Creatinine (μmol/L)	0·09 (0·07 to 0·12)	<0·0001	0·08 (0·06 to 0·10)	<0·0001
Folate (μg/L)	–0·23 (–0·36 to –0·10)	0·001	–0·15 (–0·28 to -0·03)	0·02
B12 (ng/L)	–0·003 (–0·01 to 0·001)	0·12	..	..
Holotranscobalamin (pmol/L)	..	..	–0·04 (–0·06 to –0·02)	0·001

Generalised linear models (GLMs) of potential tHcy determinants showing estimates and strengths of their respective effects

Gene-expression profiling and identification of patients at high risk of breast cancer

Sir—André Ahr and colleagues (Jan 12, p 131)[1] report on the use of cDNA microarray and cluster analysis of gene-expression patterns to prospectively identify a subset of patients with primary breast cancer at high risk of subsequent disease recurrence.

In their initial report,[2] tumours were split into four main groups, namely I–IV, on the basis of cluster analysis.[3] Group III was further separated into class A and class B. Class A was described as containing transcripitionally related samples that were further separated into two subpopulations designated A1 and A2 based on the differences in transcription of several genes. A2 has an exceptionally low expression of oestrogen receptor α and progesterone receptor, *BAD* gene, and insulin-like growth factor binding protein 2, compared with A1.

In the first report, Ahr and colleagues analysed and presented the data on A1 alone, and presented the data on A2 along with the tumours in the other groups, since the A1 subpopulation was characterised by a disproportionately high frequency of lymph-node-positive tumours and distant metastasis at the time of diagnosis compared with the other tumours.

However in the follow-up analysis of these patients,[1] the investigators no longer refer to the molecular subdivision in class A and the differences previously noted between A1 and A2 for lymph-node status and distant metastasis; they simply refer to patients as being class A and non-class A.

Given that the molecular differences between A1 and A2 tumours are small, if the clinical behaviour of these two subpopulations continues to differ it could provide important clues as to the gene or genes that may best predict the risk of disease recurrence and metastasis in breast cancer. Therefore, the presentation of data on the subpopulations A1 and A2 in the follow-up analysis is important.

*Carlo Palmieri, David Vigushin

Department of Cancer Medicine, Cancer Research-UK Laboratories, Faculty of Medicine, Imperial College School of Science, Technology and Medicine, Hammersmith Hospital, London W12 0NN, UK
(e-mail: c.palmieri@ic.ac.uk)

1 Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 2002; **359:** 131–32.
2 Ahr A, Holtrich U, Solbach C, et al. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol* 2002; **195:** 312–20.
3 Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95:** 14863–68.

Sir—Emerging high-throughput molecular analyses have the potential to reveal previously unknown prognostic subclasses of tumours with more homogeneous clinical outcome. By use of DNA array technology, André Ahr and colleagues[1] identify a class of breast-cancer patients with a high risk of metastasis.

These results are clinically important since progress in breast cancer therapy is expected to come as much from the

development of specific antitumour drugs as from the establishment of a prognostic or predictive classification that would reflect the heterogeneity of disease more accurately than current histoclinical factors. Such a classification might lead to better tumour-tailored treatments. However, in their study, several issues in the methods potentially diminish, in our opinion, the impact of the results.

Ahr and colleagues exclude T4 tumours and details about treatment from the survival analysis, and other variables relevant for prognosis (patients' age and hormonal-receptor status, and histological type and grade of tumours) are not provided. Furthermore, the discriminator gene set they select is based on their differential expression between normal and cancerous mammary tissues.[2] These so-called diagnostic genes also seem relevant for prognosis.

A rapid way to assess and increase the validity of the present results is to confront them with those of similar independent studies. Three studies have directly addressed the prognostic issue of breast cancer by analysing series of tumours homogeneous at histoclinical and treatment level.[3–5] By monitoring the expression of candidate genes in poor-prognosis primary breast cancers treated with adjuvant chemotherapy, we identified subgroups with different clinical outcome.[3] Similarly, gene-expression profiles revealed new prognostic subclasses in patients with primary breast cancers that had good prognosis receiving no adjuvant therapy,[5] and with locally advanced breast cancers treated with neoadjuvant doxorubicin.[4]

We compared the discriminator genes between these four studies with an extensive cross-analysis. Despite the use of different approaches, we noted several genes that were present in at least two studies. Some are associated with prognosis (*ERBB2, ESRI*), which validates the approach and the study of Ahr and colleagues. On the other hand, most are not currently known as predictive factors (*CP*, *GATA3*, *IGFBPs*, *OXCT*), and display a broad range of functions, such as cell proliferation, apoptosis, invasion, angiogenesis, kinases, or transcription factors, which might make them new molecular targets for anticancer treatments.

All these results are hence promising for the future, and constitute a starting point for innovative tackling of breast tumours. But before any clinical application is routinely implemented, the benefits for patients will have to be shown. In this context, the required and expected public availability of DNA arrays data will allow computational comparison and validation of results to boost and strengthen the analyses.

*François Bertucci, François Eisinger, Rémi Houlgatte, Patrice Viens, *Daniel Birnbaum*

Departments of *Molecular Oncology, Prevention, and Medical Oncology, Institut Paoli-Calmettes, INSERM U 119, 13009 Marseille, France; TAGC laboratory, Marseille; and University of Méditerranée, Marseille (e-mail: birnbaum@marseille.inserm.fr)

1 Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 2002; **359:** 131–32.
2 Ahr A, Holtrich U, Solbach C, et al. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol* 2001; **195:** 312–20.
3 Bertucci F, Houlgatte R, Benziane A, et al. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet* 2000; **9:** 2981–91.
4 Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; **98:** 10869–74.
5 van 't Veer LJ, Dai H, van De Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415:** 530–36.

Authors' reply

Sir—Carlo Palmieri and David Vigushin ask about the subgroups in our high-risk class of mammary carcinomas.

To clarify, in the first report we used a hierarchical clustering algorithm on gene expression data for the class discovery of breast tumours resulting in the main classes A, B, and non-AB. Class A, which defines the high-risk cancer population, could be further separated into subclasses A1 and A2.

For verification of this clustering, we used the method of class prediction, leave-one-out cross validation.[1] The resulting prediction strength for each sample, a value between 0 and 1, gives a measurement of how safely a sample can be assigned to the corresponding group. Comparison of all samples in class A with non-A samples gave a median prediction strength of 0·727. By contrast, the distinction between A1 and A2 was weak (0·54). Furthermore we used self-organising maps[2] as an independent method to classify the tumour samples.[3] The self-organising-map algorithm was able to reliably identify classes A and non-A, when a two-class self-organising map was applied (http://www.kgu.de/zfg/dnachip [accessed April, 2002]). However, the A1/A2 distinction was not predicted by this method.

Although the subdivision of class A is not stringently predicted by our marker set, the original identified class A2 differs from A1 by a strong down-regulation of the oestrogen α and progesterone receptors. When we compared the rate of recurrences between those two subgroups, as Palmieri and Vigushin request, it was slightly higher in A2 than in A1 (three of six, 50% *vs* six of 16, 38%). Two of the three recurrences observed in A2 were lymph-node-negative patients. These data might suggest a raised risk for A2 patients. Thus, although the number of patients in A2 is too small for a significant analysis, identification of additional marker genes that can discriminate more definitely between the two subgroups may be important.

François Bertucci and colleagues criticise some features of our methods. We excluded T4 tumours since tumours of this type represent a priori a high-risk group. Only two of eight T4 tumours in our patients showed no recurrences during follow-up. These two cases were classified as non-A, in line with the lower risk in this class. Second, we analysed potential correlations of our classification with standard variables such as patients' age, histological type, and grading of tumours. We noted no significant correlation.

The statement that our gene selection was based on differential expression between normal and cancerous tissues is not correct. We selected the gene set on the basis of differential expression between different mammary carcinomas. Purified normal mammary epithelial cells were used only in the process of relative standardisation for comparison of different arrays. We agree with Bertucci and colleagues that further standardisation of microarray data will lead to better comparability and accelerate the development of clinical applications.

**André Ahr, Thomas Karn, Uwe Holtrich, Manfred Kaufmann*

Department of Obstetrics and Gynaecology, J W Goethe-University, 60590 Frankfurt, Germany (e-mail: Ahr@em.uni-frankfurt.de)

1 Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286:** 531–37.
2 Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999; **96:** 2907–12.
3 Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 2002; **359:** 131–32.