Gene expression profiling in single cells within tissue

Paola Capodieci[1], Michael Donovan[1], Heidi Buchinsky[1], Yusuf Jeffers[1], Carlos Cordon-Cardo[2], William Gerald[2], Jon Edelson[1], Shailesh M Shenoy[3] & Robert H Singer[3]

We developed a robust multiplex fluorescent *in situ* hybridization (FISH) technique in archival formalin-fixed, paraffin-embedded (FFPE) human tissue sections while preserving the microanatomical context. This identifies single-cell gene expression patterns by probing multiple, unique nascent RNA transcripts and yields predictive quantitative gene expression signatures.

Integrating histological features and molecular profiles enhances our understanding of cellular response to the microenvironment and elucidates biological mechanisms of disease processes. This integration has the potential to improve patient management, assisting in early diagnosis and objective patient prognosis, including outcome and response to treatment. As an example, we investigated gene expression signatures in prostate tumor and benign cells from individual patient samples.

Most gene expression analyses, unlike FISH studies, extract RNA from solubilized tissue, which destroys cellular architecture such that expression profiles cannot be associated with specific cell types. Analysis of small samples excludes the detection of many potentially interesting low-abundance gene products. This coupled with the lack of standardization and reproducibility as well as amplification bias has limited the interpretation of expression data resulting from methods such as RNA amplification and quantitative PCR[1–3].

Development of FISH protocols using cultured cells yields single-cell expression profiles previously unobtainable by other methods[4,5]. The approach detects signal foci from multiple nascent RNA localized at the transcription site of genes that are expressed. The presence of nascent RNA directly correlates with early events of gene regulation in response to external or physiological stimuli[6]. This provides an opportunity to quantitatively evaluate the temporal expression pattern of several genes in single nuclei.

We describe a new paraffin-embedded tissue FISH (peT-FISH) method to simultaneously detect expression of several genes *in situ* in single cells while maintaining tissue morphology. The method uses automated fluorescence microscopy and analysis to detect nascent RNA in archival tissue specimens from both whole-section or tissue-microarray (TMA) format, ultimately allowing high-throughput gene expression analysis that can be linked to clinical outcome data.

The antigen retrieval method that uses sodium borohydride and high heat[7] (**Supplementary Methods** online) preserves the integrity of nascent RNA, promotes robust hybridization and reduces autofluorescence, allowing the detection of genes expressed at a low level (for example, *JAG1*). Housekeeping genes such as *ACTB* and

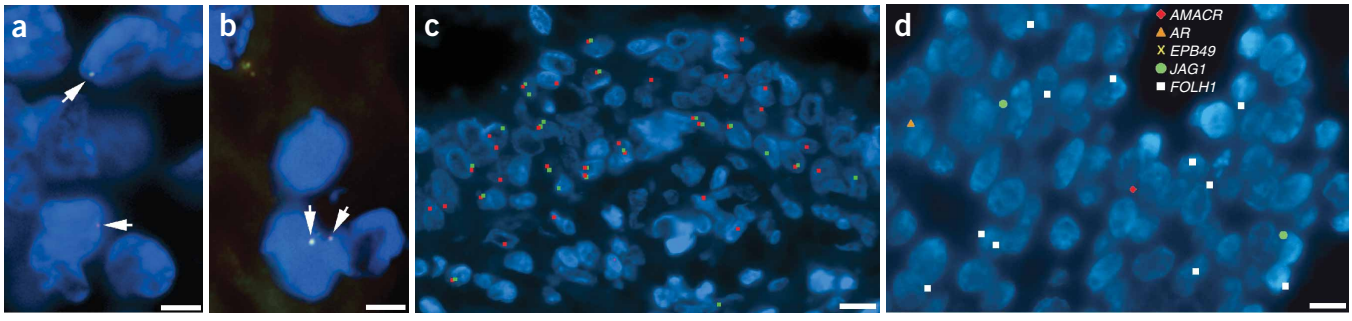

Figure 1 | Detection of nascent RNA (transcription sites) in paraffin-embedded tissue. (**a**) Detection of *SMG1* (8 probes) transcription sites (arrows) in human PCa. Bar, 5 μm. (**b**) Detection of *SMG1* (41 probes) nascent transcripts (arrows) in human PCa. Bar, 5 μm. (**c**) Colocalization of a gene and the respective chromosome. Prostate carcinoma exhibiting the colocalization of a DNA locus–specific probe for *AR* (androgen receptor, red spots) with the *AR* peT-FISH on a single section (green spots). Bar, 10 μm. (**d**) Multiplex detection of five genes in a prostate cancer sample (1982 archive), demonstrating restricted expression of four genes to the epithelial cells of the tumor annotated with the automated transcription site finder algorithm. The fifth gene in the barcoding scheme, *EPB49*, was only rarely detected and was not identified in this particular field. Bar, 10 μm.

[1]Aureon Laboratories, 28 Wells Avenue, Yonkers, New York 10701, USA. [2]Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA. [3]Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. Correspondence should be addressed to P.C. (paola.capodieci@aureon.com) or R.H.S. (rhsinger@aecom.yu.edu).

PUBLISHED ONLINE 23 AUGUST 2005; DOI:10.1038/NMETH786

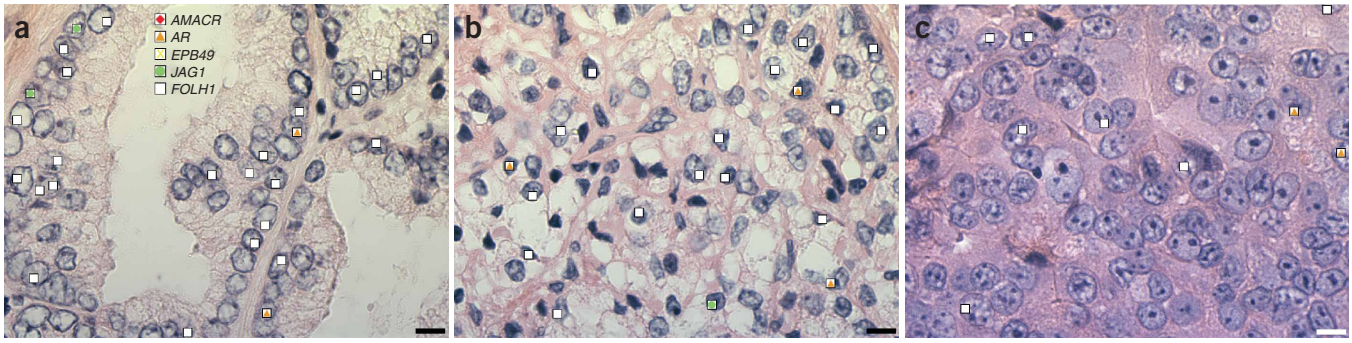

Figure 2 | Correlation of single-cell multigene expression profiles with diagnostic pathology. (**a**–**c**) Hematoxylin and eosin–stained images of focal PIN (**a**), PCa (**b**) and PCaMet (**c**), superimposed with the gene expression data from peT-FISH analyses of three genes (*AR*, *JAG1* and *FOLH1*) obtained from the slide after destaining. Bars, 10 μm.

SMG1, a phosphatidylinositol 3 kinase–related protein kinase, were used for quality control standards. The number of *SMG1* transcription sites detected by peT-FISH was unchanged when the number of oligonucleotide probes targeting *SMG1* ranged from 8 to 82 (**Fig. 1a**,**b**); however, six probes detected 50% fewer transcription sites (data not shown). This indicated the optimal number of probes required to detect transcription reproducibly and robustly. Simultaneous peT-FISH and DNA FISH confirmed the specificity of transcription site detection, as signal from nascent RNA and the gene were coincident (**Fig. 1c**). Signal from DNA FISH targeting a different gene did not correlate with the peT-FISH signal (data not shown). Detection of nascent transcripts was not restricted to freshly prepared tissue samples. This method robustly detected expression of five genes from archived FFPE tissue samples, of which the majority were 10–15 years old and one was 22 years old (**Fig. 1d**).

To investigate how peT-FISH data correlate with histopathological features, we examined epithelial cells with established pathologic morphologies: benign prostate tissue, prostatic intraepithelial neoplasia (PIN), prostate cancer (PCa) and prostate cancer metastasis (PCaMet)[8]. peT-FISH was performed on destained hematoxylin and eosin tissue sections (**Fig. 2**), an important capability to have when patient tissue blocks are no longer available. Overlays of peT-FISH results on morphology image data are shown. The procedure was modified for high-throughput sampling of patient material in TMA format[7].

Software detected and analyzed transcription sites and decoded each gene's identity using the multiplexed barcode (**Supplementary Methods**). The correlation of peT-FISH data with morphological features in a PCa progression TMA from 59 patients indicated that expression patterns allow for the identification and subsetting of tumor-specific cells in all of the patient samples (**Fig. 3** and **Supplementary Tables 1–3** online). The transcriptional profile of *AMACR*, *AR*, *JAG1* and *FOLH1* in PCa samples was significantly different from the profile from benign prostate tissue, focal PIN and PCaMet, yielding statistically significant, disease-specific combinatorial differences (**Supplementary Table 3**). Unique gene expression 'fingerprints' manifested discrete classes of patient samples for similar histopathological types, suggesting true biological similarities. In contrast to the epithelium, we did not identify a signal in the nuclei of the noncancerous stromal fibroblasts or endothelial cells. peT-FISH technology facilitates the development of integrated molecular and cellular roadmaps of clinical disease. We now have the capability to correlate morphological features with the molecular basis of pathogenesis.

Microarray and quantitative PCR technology provide refined disease classification schemes based on gene expression patterns rather than classic morphology[9,10] resulting from correlation of multigene expression patterns from archived tissue with a clinical outcome, a response to therapy or a physiologic event[11,12]. These technologies cannot assign gene expression status to specific cell types and cannot reliably detect low or transiently expressed genes. peT-FISH is the only technology that measures *in situ* multiplex gene expression while retaining tissue morphology. The published association between gene expression profiles and changes in cellular morphology in human tissue samples is rare. Recently the transcriptional coactivator p300, previously linked to prostate cancer progression, and the nuclear lamins were found to modulate

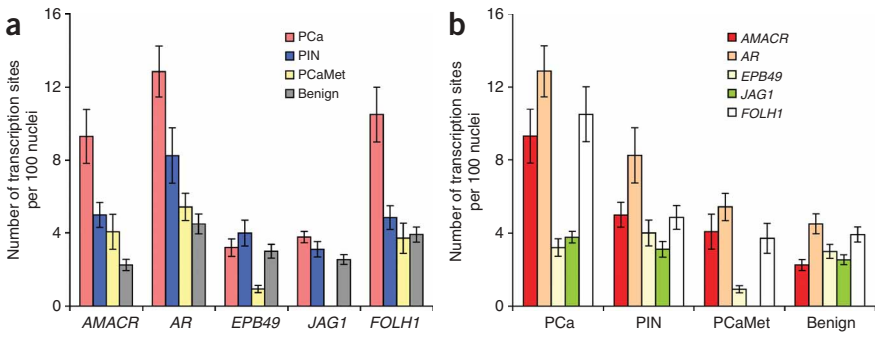

Figure 3 | A combinatorial analysis of five genes provides a gene expression signature for each state. (**a**) The gene expression patterns of each diagnosis as a function of each gene (*AMACR*, *AR*, *EPB49*, *JAG1* and *FOLH1*). (**b**) The gene expression patterns characteristic of each diagnosis (focal PIN, PCa, PCaMet and benign). Data represent the mean ± s.e.m.; $n = 20$, 13, 14 and 12 for PCa, PIN, PCaMet and benign tissue, respectively.

nuclear morphology in prostate cancer epithelial cells[13]. Our peT-FISH data support a link between morphology and gene expression profiles, and thus have the potential to modify present-day tumor grading systems or identify differentiation states.

peT-FISH can potentially detect noncoding genes such as microRNAs and certain transcription factors for which the necessary reagents to establish cellular localization are lacking[14]. The noncoding gene, *PCA3* (also known as *DD3*), is an example where selective association with prostate cancer epithelial cells has made this a unique target for both the diagnosis and treatment of prostate cancer[15]. peT-FISH technology enhances the biological knowledge for a given gene by associating its expression to a specific cell type (that is, endothelial versus epithelial versus inflammatory versus stromal cells) and through comparative quantitative assessment with multiple genes, can establish a link with various signaling cascades. One may anticipate that these additional correlative data will aid in reaching the ultimate goal—identifying and using key signatures, based on association with clinical outcome data, to establish predictive and prognostic markers. We expect the enormous information inherent in the expression of many genes in large cell populations will aid the understanding of relationships among genes in single nuclei and their cooperative and cumulative roles in physiology and disease.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported in part by US National Institutes of Health CA-R33-083208 and by Aureon Laboratories.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 25 April; accepted 25 July 2005
Published online at http://www.nature.com/naturemethods/

1. Lockhart, D.J. & Winzeler, E.A. *Nature* **405**, 827–836 (2000).
2. Enders, G. *Acta Neurochir.* (Suppl.) **89**, 9–13 (2004).
3. Hoffmann, K., Firth, M.J., Freitas, J.R., de Klerk, N.H. & Kees, U.R. *Mol. Biotechnol.* **29**, 31–38 (2005).
4. Levsky, J.M. *et al. Science* **297**, 836–840 (2002).
5. Femino, A.M. *et al. Science* **280**, 585–590 (1998).
6. Wilson, A.J. *et al. Cancer Res.* **62**, 6006–6010 (2002).
7. Capodieci, P. *et al. J. Natl. Cancer Inst.* **96**, 269–279 (2004).
8. Glinsky, G. *et al. J. Clin. Invest.* **113**, 913–923 (2004).
9. Heighway, J. *et al. Oncogene* **21**, 7749–7763 (2002).
10. Cronin, M. *et al. Amer. J. Pathol.* **164**, 35–42 (2004).
11. Paik, S. *et al. NEJM* **351**, 2817–2826 (2004).
12. van de Vijver, M.J. *et al. NEJM* **347**, 1999–2009 (2002).
13. Debes, J.D., *et al. Cancer Res.* **65**, 708–712 (2005).
14. Lu, J. *et al. Nature* **435**, 834–838 (2005).
15. Bussemakers, M.J. *et al. Cancer Res.* **59**, 5975–5979 (1999).