

Amplification-free digital gene expression profiling from minute cell quantities

Fatih Ozsolak^{1,5}, David T Ting^{2,3,5}, Ben S Wittner^{2,3}, Brian W Brannigan², Suchismita Paul², Nabeel Bardeesy^{2,3}, Sridhar Ramaswamy^{2,3}, Patrice M Milos¹ & Daniel A Haber²⁻⁴

Generating reliable expression profiles from minute cell quantities is critical for scientific discovery and potential clinical applications. Here we present low-quantity digital gene expression (LQ-DGE), an amplification-free approach involving capture of poly(A)⁺ RNAs from cellular lysates onto poly(dT)-coated sequencing surfaces, followed by on-surface reverse transcription and sequencing. We applied LQ-DGE to profile malignant and nonmalignant mouse and human cells, demonstrating its quantitative power and potential applicability to archival specimens.

Recent advances in both cancer and stem cell biology have highlighted the biological importance of extraordinarily rare cells that may have fundamentally different gene expression patterns than surrounding cell types. Although next-generation sequencing technologies bring unprecedented power to gene expression studies, the inability to apply reliable molecular profiling analyses to such minimal numbers of cells has proven to be a major limitation of current sequencing-based strategies. Methods developed to date rely on multiple sample manipulation and amplification steps, which introduce errors and skew the original representation of the nucleic acid population¹⁻⁴, rendering these methods unsuitable for applications requiring high fidelity.

To profile minute RNA quantities for gene expression patterns, we extended the single-molecule sequencing (SMS) technology^{5,6} to establish a low-quantity digital gene expression (LQ-DGE) application involving direct flow-cell capture of RNA. To optimize and test LQ-DGE, we selected two related yet distinct cell types: SM25 and 490. SM25 has been derived from a pancreatic intraepithelial neoplasia lesion in a genetically engineered mouse with pancreas-specific expression of the *Kras*^{G12D} mutant⁷. Cell line 490 has been established from a malignant pancreatic ductal adenocarcinoma (PDAC) lesion, combining the *Kras*^{G12D}

mutation with conditional loss of the *Trp53* gene⁸. Comparison of these two cell lines made it possible to model biologically important genetic differences among related cell types.

We performed LQ-DGE using 250 to 16,000 cells. Briefly, we captured poly(A)⁺ mRNA from cell lysates on the poly(dT)-coated sequencing flow cells. We initiated on-surface cDNA synthesis using SuperScript III reverse transcriptase, followed by terminal transferase-mediated on-surface guanine (G)-tailing of cDNAs covalently attached to surfaces to generate priming sites allowing sequencing from the 'top' (cDNA 3' ends) (Fig. 1). We hybridized a poly(C) primer to the G-tailed templates, followed by a 'fill-and-lock' step⁶. Then we initiated SMS without additional modifications.

To determine the effect of cell quantity on the number of usable reads obtained per channel, we performed a titration experiment (Fig. 2a). As few as 250 cells generated sufficient usable reads for digital gene expression (DGE) profiling (Supplementary Fig. 1). The measurements were highly reproducible, as demonstrated by profiling 1,000 cells, using lysates prepared at separate times in two independent runs (Fig. 2b). Our comparison of four different commercial cell lysis conditions showed a high correlation between transcript counts obtained ($r = 0.937-0.946$; Supplementary Fig. 2). However, these correlations were lower than the correlation obtained from profiles generated with the same lysis condition, suggesting that the LQ-DGE profiles may be slightly dependent on cell-lysis methods.

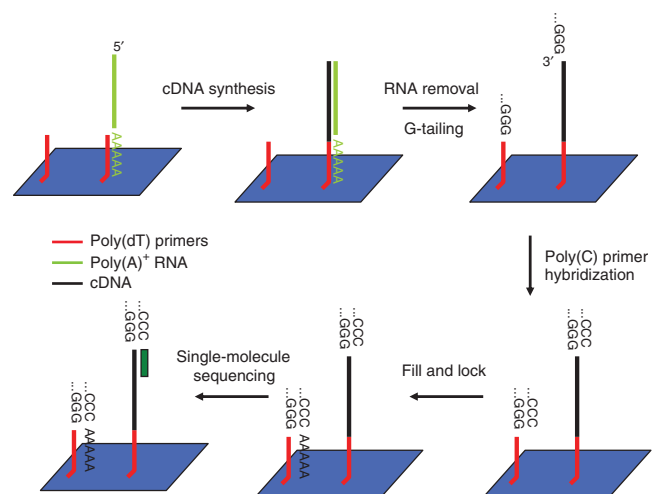


Figure 1 | LQ-DGE template capture, on-surface cDNA synthesis and sequencing workflow. After capturing the poly(A)⁺ RNA on surfaces coated with covalently attached poly(dT) primers, natural dNTPs and reverse transcriptases are used to synthesize cDNA. Indicated steps are followed to obtain cDNA sequence by synthesis (green rectangle).

¹Helicos BioSciences Corporation, Cambridge, Massachusetts, USA. ²Massachusetts General Hospital Cancer Center, Charlestown, Massachusetts, USA. ³Harvard Medical School, Boston, Massachusetts, USA. ⁴Howard Hughes Medical Institute, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts, USA. ⁵These authors contributed equally to this work. Correspondence should be addressed to F.O. (fatihozsolak@gmail.com) or D.A.H. (haber@helix.mgh.harvard.edu).

