

Identifying pattern in microarray expression series using algorithmic information theory

Ahnert SE¹, Brown FCS², Fink TMA³, Willbrand K⁴

¹*Theory of Condensed Matter, Cavendish Laboratory, Cambridge CB3 0HE, UK*

²*DMA, Ecole Normale Supérieure, 75231 Paris Cedex 05, France*

³*Institut Curie, CNRS UMR 144, 75248 Paris Cedex 05, France*

⁴*Laboratoire de Physique Statistique, Ecole Normale Supérieure, 75231 Paris Cedex 05, France*

karen.willbrand@medizin.uni-essen.de

Abstract We introduce a method of detecting pattern in data series independent of the nature of the pattern. Our approach is to replace each data series with an alternative description from which the original data can be fully recovered. Data series with short descriptions, which are significantly compressible are more likely to result from simple underlying mechanisms than series which are incompressible. We assume compressible gene expression profiles to be biologically or medically interesting. We show that the compression in bits k is a universal currency which is independent of the type of noise underlying the data and by which we can order data series according to their significance. The method is successfully tested on microarray time series of yeast cell cycle.