Gene Expression Markers of Proliferation and Differentiation in Cancer and the Extent of Prognostic Signals in the Cancer Transcriptome

Gil Tomás

January 10, 2014

Contents

1 Introduction

1.1 Life

A brief recapitulation on the canonical perspective of the evolutionary history of life, starting from the central dogma of mollecular biology and then from unicellular to multicellular levels of integration and complexity. Introduces the concepts of evolution and adaptation, homeostasis (homeodynamics), developmental biology, high level structural tissular integration, response to stimuli and cell-to-cell communication.

This section of the introduction will provide the reader with some of the core concepts around which the discussion will be framed.

1.2 Cancer

Cancer is first portrayed as a by-product of the complexity of biological systems resulting in a disruption of homeostatic processes. Inasmuch, a first conceptual motivation for a better understanding of cancer is a further elucidation of the mechanisms governing the normal functioning of multicellular life. Cancer is thus primarily introduced as departure from sustained homeodynamic equilibrium in organisms.

1.2.1 Cancer: a clinical perspective

Study of cancer is then motivated from a clinical perspective and discussed as a global public health issue. This section will cover the epidemiology of cancer, possible causes, generic methods of diagnosis and treatment.

Incidence of cancers A summary of the most recent WHO cancer incidence data will be presented here.

1.2.2 The biology of cancer

This section reviews the biological capabilities acquired during the multistep development of human tumors. It frames cancer as a disease of genomic instability and discusses its mollecular origin and evolutionary dynamics.

Discusses the major mollecular pathways linked with cancer. Introduces the concepts of cell transformation, apoptosis, necrosis, tumour angiogenesis, epithelial-to-mesenchymal transition (EMT), immune detection, invasion and metastasis.

1.2.3 Differentiation in cancer

Once transformed, cancer cells incur in a gradual deregulation of the patterns of gene expression associated with their tissue of origin.

Discusses the concept of *grade* in cytopathology and the clinical, diagnostic and mollecular implications of dedifferentiation in cancer.

1.2.4 Proliferation in cancer

Cancer is a disease of uncontrolled cell division and thus proliferation is a prime hallmark of cancer.

Discusses the clinical assessments of proliferation in cancer and their mollecular underpinnings.

1.3 Microarrays

1.3.1 A short history

Provides a short introduction to microarrays starting from the early 1990's as an evolution from Southern blotting to measure genome-wide gene expression.

Discusses what are microarrays; two-channel *vs* one-channel detection; experimental designs.

1.3.2 Microarrays and cancer

Provides the reader with a perspective on the impact of microarray technology in the cancer research field. Discusses the prospectives of technology raised in the early 2000's. Some discussed major achievements include:

- discovery of cancer-specific mollecular subtypes and their relationships with previously described histological subtypes (breast and medullobalstoma examples);
- use of microarrays in diagnosis and prognosis of cancer;
- insights on eventual biological mechanisms driving cancer.

Introduces the reader to survival analysis.

1.3.3 Microarray data analysis and bioinformatics

Discusses the analytical bottlenecks raised by the sheer volume of data genereated by microarray technology. Covers the concepts of class discovery & class prediction analysis; hipothesis-driven statistical analysis and discusses the relation between probe and gene. Introduces the Gene Expression Omnibus.

1.4 Motivation and main contributions of this thesis

1.4.1 Main goals

The motivation of this thesis is two-fold:

1. To explore the use of differentiation metagenes as markers of the progression of cancers, both from the perspective of their loss of differentiation as well as of their concomitant increased aggressivity.
2. To dissect the genetic programs contributing to the prognostic content of expression profiles of patients of distinct types of cancer, as assessed by survival analysis. To ascertain the impact of proliferation genetic programes in the extent of these prognostic signals.

This research program was carried making use of publicly available gene expression profiles, open source software and by implementing data analysis methodologies in reproducible computational routines. The findings obtained during this thesis are then examined in the discussion section from the perspective of the use of microarray technology both to diagnose and prognosticate cancer, as well as acquire insight in the biology of cancer.

2 Publications

1. Principal papers

(a) oncogene paper

(b) breast cancer cohorts paper, to be submitted

2. Collaborative papers

(a) 5-aza

(b) Epac

3 Material and methods

3.1 Data analysis

Discusses from a conceptual point of view: supervised *vs* unsupervised analysis; challenges associated with of high-throughput technology outputs; open source software for data analysis

Key concepts: bioinformatics; trustworthy software (Prime Directive); reproducible research and its challenges

3.2 Data collection

A description of the Gene Expression Onmibus and other online repositories of microarray data resources.

3.3 Microarray analysis

Introduces from a technical point of view: data preprocessing; dimensionality reduction; clustering; heatmap; principal component analysis; censoring data and survival analysis

4 Results

4.1 Methods to extract differentiation signatures

We developed tools to derive gene expression signatures (metagenes) from expression profiles of healthy tissues.

4.2 Differentiation signatures have diagnostic potential in cancer

We showed that such metagenes are able to discriminate between cancer subtypes of distinct agressivity.

4.3 Prognostic content is heterogeneous across distinct cohorts of breast cancer

We highlighted the heterogeneity of the prognostic content of distinct cohorts of breast cancer expression profiles and characterized the range of biological and technical variables that could account for it.

4.4 Proliferation captures most of the prognostic content of cancer transcriptomes

We reassessed the extent by which proliferation genetic programs could impact global expression profiles in breast cancer and therefore account for the majority of prognostic signals therein.

Focus on introducing the main results of each paper in a sequential way:

- motivation (question addressed)
- methods (data analysis techniques employed)
- results (graphical or tabular output)
- findings (brief description of conclusions drawn)

Reviews supplementary data of each paper in the same terms. PDF copies go in appendix.

5 Discussion

5.1 Biological insights

Discusses the contributions of this thesis from a biological point of view, namely:

- cancer as a reversion of a differentiation program (examines cancer stem cells);
- an eventual pan-transcriptomic impact of genetic programs linked to cell dievision and proliferation;
- how to discretize transcriptomic signals from distinct cellular types (*e.g.*, how to quantify immune response of inflammatory response with microarrays)

5.2 Technical insights

Discusses the contributions of this thesis from the technical point of view, namely:

- strengths and limitations of microarrays
- strengths and limitations of methodological statistical tools employed
- eventual ways of addressing technological and methodological limitations

5.3 Impact of microarrays on cancer research

Discusses how microarray technology impacted and reshaped cancer research, from the diagnosis to prognosis; from enhancing our understanding of the pathways disrupted in cancer to validation of new therapeutic targets.

Confronts the expectations raised by the technology the early 2000's with the current state-of-the-art understanding of cancer, now at the dawn of another technological wave. Concludes with an epistemological note on the scientific process (the decline effect, Ioannidis' research)

5.4 Prospects in cancer research

Provides a brief account of new mainstream technologies employed in cancer research, including next-generation sequencing and bead arrays. These technologies have largely replaced microarray technology as the prime tool to investigate gene expression patterns in cancer. Discusses how the lessons drawn from 15 years of microarray-driven cancer research and the insights discussed in this thesis may benefit the

6 Conclusion

Wraps up with a broader perspective on cancer by framing its progression through the perspective of evolutionary life history. Covers concepts such as cellular Darwinism, tumour heterogeneity, tumour dormancy and aggressivity. Proposes a reinterpretation of the findings of this thesis through this point of view, contextualized by the limitations of microarray technology discussed above.