Chapter 18

Ribonoscopy and Personalized Medicine

Excerpt from


As pointed out in Sect. 12.6, microarrays can be used to measure either DNA or RNA from cell samples (see Fig. 12.5). Hence, it may be useful to coin two words, ribonoscopy and deoxy-ribonoscopy to distinguish between these two types of measurements. Confusing these two different measurements can lead to logical errors in interpreting microarray data (Ji et al. 2009a). The term “ribonoscopy” is composed of two stems – “ribo” meaning ribonucleic acid or RNA, and “-scope” meaning to look carefully or to see – and hence “ribonoscopy” literally means “to carefully look at RNAs” (rather than genes or DNA) with respect to the changes in both their sequences (or kinds, quality) and abundances (also called levels or concentrations, quantity). More formally ribonoscopy can be defined as

\[
\text{The study of the genome-wide RNA sequences and concentrations inside the cell measured with cDNA microarrays and visualized as spectra (called ribonic spectra) with the y-axis registering RNA levels and the x-axis recording time, structure, or other related variables.}
\]

\[(18.1)\]

It is often useful to distinguish between the raw, unprocessed microarray data and the results of the analysis of the raw data using software such as hierarchical clustering (Eisen et al. 1998) and ViDaExpert (Gorban and Zinovyev 2004, 2005). The former (i.e., RNA trajectories) is referred to as “ribons” (see Sects. 12.8.2 and 12.8.3) and the latter (i.e., 2- or 3-D displays of the results of analysing microarray data) as “ribonic spectra” or “ribospectra.” The relation among ribons, computer-assisted analytical tools, and ribonic spectra can be illustrated using an analogy between optical spectra in physics and ribonic spectra as shown in Fig. 18.1. Thus ribons are akin to light, computer softwares are analogous to a prism, and ribonic spectra are comparable to optical spectra.

The content of ribonoscopy is explained in a greater detail in Table 18.1. There are two major types of ribons – (1) the t-ribons, namely, time series as exemplified by RNA trajectories (e.g., Fig. 9.1) and (2) what is here referred to as the s-ribons, or structure series (see Table 9.1 and Figs. 19.2 and 19.3) in analogy to “time series.” These two types of ribons are the inputs to ribonoscopy as indicated in the first column of Table 18.1. The raw data from microarray measurements on n RNAs
can be represented as a set of $n$ points in an $N$-dimensional mathematical space (to be called the “ribonic space” or “RNA concentration space”), where $N$ is either the time points of measurements or the number of different samples analyzed. The ribonic space is depicted in the top figure in the second column of Table 18.1, where the undulating membrane indicates a principal manifold onto which the nearest points are projected by the ViDaExpert program (Sect. 2.8.1). The ViDaExpert software is based on several well-established mathematical and computational frameworks. Each of the $n$ point in the $N$-dimensional concentration space represents a ribon with $N$ coordinate values which is the main reason for referring to the mathematical space as the ribonic space. In the case of a $t$-ribon (e.g., see the upper figure in the first column of Table 18.1), its coordinate values are the levels (or copy numbers) of an RNA measured at $N$ different time points. In the case of an $s$-ribon (e.g., see the lower figure in the first column of Table 18.1), its coordinate values represent the levels or copy numbers of an RNA measured in $N$ different samples.

The raw data of microarray measurements can also be represented as a “distance matrix” (see the table in the second column of Table 18.1) where the Euclidean distances between all possible pairs of the points (i.e., ribons) in the $N$-dimensional concentration space have been calculated based on the Pythagorean formula. It should be noted that

\[
\begin{align*}
\text{All the information contained in the raw RNA data measured with microarrays is encoded} \\
\text{in the distance matrix which is symmetric with respect to its diagonal because the distance} \\
\text{between a and b is the same as that between b and a.} \\
\end{align*}
\]

(18.2)

The distance matrix defined in Statement 18.2 may be referred to as the “ribonic matrix,” since the first row and the first column of the matrix are composed of the names of the ribons measured by microarrays.

It is possible to analyze the raw data, that is, both $t$- and $s$-ribons, in the forms of ribonic matrices, without relying on any specialized analytical soft wares such as hierarchical clustering or ViDaExpert. The top two figures in the third column of Table 18.1 represent significant results of analyzing ribonic matrices without utilizing any computational softwares (Ji et al. 2009b, c). The third figure in the
Table 18.1 Various forms of inputs, methods of analysis, and outputs underlying ribonoscopy, the study of RNA molecules with respect to their sequences and copy numbers in the living cell

<table>
<thead>
<tr>
<th>Input (RNA raw data or &quot;Ribons&quot;)</th>
<th>Computer-aided analysis</th>
<th>Output (ribonic spectra)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time-Series Data</strong> (e.g., RNA trajectories)</td>
<td>1. N-dimensional Ribonic space</td>
<td><img src="image1" alt="Graphic" /></td>
</tr>
<tr>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Matrix" /></td>
<td>2. Ribonic matrix</td>
</tr>
<tr>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Table" /></td>
<td>a</td>
</tr>
<tr>
<td><img src="image6" alt="Graph" /></td>
<td>3. Genotype-phenotype correlation analysis</td>
<td>a</td>
</tr>
<tr>
<td><img src="image7" alt="Graph" /></td>
<td>4. Histograms</td>
<td>b</td>
</tr>
<tr>
<td><img src="image8" alt="Graph" /></td>
<td>5. Clustering</td>
<td>c</td>
</tr>
<tr>
<td><img src="image9" alt="Graph" /></td>
<td>6. ViDaExpert</td>
<td>d</td>
</tr>
<tr>
<td><strong>Structure Series Data</strong> (e.g., Gene expression profiles)</td>
<td></td>
<td><img src="image10" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image11" alt="Graph" /></td>
<td></td>
<td><img src="image12" alt="Gene expression" /></td>
</tr>
<tr>
<td><img src="image13" alt="Graph" /></td>
<td></td>
<td><img src="image14" alt="Gene expression" /></td>
</tr>
<tr>
<td><img src="image15" alt="Graph" /></td>
<td></td>
<td><img src="image16" alt="Gene expression" /></td>
</tr>
</tbody>
</table>

*Note: Graphs and tables are placeholders for actual visual representations.*
third column of Table 18.1 is the result of clustering the ribonic matrices measured from human breast tissues and tumors (Perou et al. 2000). The fourth and fifth figures in the third column of the same table are the 3-D and 2-D visualizations of the ViDaExpert-analyzed results of the t-ribons measured from budding yeast undergoing glucose-galactose shift (Sects. 12.8.2 and 12.8.3).

One of the major assumptions of this section is that, to apply ribonoscopy to personalized medicine, it is necessary to utilize the molecular theory of the living cell such as the one developed in this book, especially the concept of dissipative structures in general and intracellular (ic) dissipative structures (IDSs or ic-dissipatons) in particular (Sect. 3.1.2). Ribonoscopy is one of the few experimental methods now available that allows IDSs or ic-dissipatons to be measured genome-wide. Personalized medicine differs from traditional medicine in that it tailors health care (through the triad of diagnosis, prognosis, and therapy) to best fit individual patients taking into account their unique genetic (i.e., nucleotide sequence-dependent) and epigenetic (i.e., non-nucleotide sequence-dependent) characteristics. The roles that ribonoscopy and the molecular theory of the living cell (MTLC) developed in this book play in personalized medicine are schematically represented in Fig. 18.2. Since the cell is the building block of the human body, it is logical to anticipate that cell biology will play a fundamental role in personalized medicine (see the top node and the bottom three nodes in Fig. 18.2).

Ribonoscopy consists of two parts – (1) the microarray data acquisition (Step 1) using cDNA microarray technology (Sect. 12.8), and (2) the dimensional reduction and visualization of high-dimensional microarray data (Step 2) in the form of what is referred to as ribonic spectra (see Fig. 12.17) using ViDaExpert or similar computer softwares. It is here assumed that, in order to analyze ribonic spectra correctly and identify the ribonic spectral characteristics reflective of a diseased cell, it is necessary to apply a comprehensive MTLC (Step 3). In other words, it is thought to be impossible to identify a biomarker from ribonic spectra without applying a comprehensive molecular model of the living cell, just as it is impossible to interpret molecular spectra without quantum mechanics, the theory of the atom. Once a correct biomarker (or a disease-related cell-state, or biomarker ribonic spectrum) is identified, it can be utilized for developing companion diagnostics (i.e., the diagnostic tools that identify the patients most likely to benefit from a drug) (Step 4), drug target discovery (Step 5), or personalized drug therapy (or pharmacotherapy) (Step 6).

There is an interesting analogy to be drawn between the nuclear power industry and drug industry on several levels as indicated in Table 18.2.

The final product of a power plant is electricity; the final product of a drug manufacturing plant is safe and efficient drugs. Both kinds of plant activities inevitably produce wastes that contribute to environmental pollution – the external environmental pollution by the nuclear power industry (e.g., the Chernobyl, Three Mile Island, and Fukushima disasters) and the internal environmental pollution by drug industry (e.g., the Vioxx fiasco). Nuclear reactor engineering that emerged in the 1940s as a spinoff from the atomic bomb production in the USA and the then-USSR is based on the theory of the atom, that is, quantum mechanics that was
**Ribonoscopy as a Tool for Personalized Medicine**

1. DNA Microarrays
2. ViDaExpert
3. Molecular Theory of the Living Cell (MTLC)
4. Biomarkers (or Disease-related Cell States)
5. (E) Companion Diagnostics
6. (F) Drug Target Discovery
7. (G) Personalized Drug Therapy

**Personalized Medicine**

![Flowchart](image)

**Fig. 18.2** A schematic representation of the roles that ribonoscopy and the molecular theory of the cell (MTLC) play in personalized medicine.

Firmly established between 1900 and 1925. In stark contrast, the theory of the cell, the building block of the human body, is still lacking. The molecular theory of the living cell developed in this book may represent one of the first, if not the first, comprehensive molecular theories of the living cell (MTLC), to the best of my knowledge. The ribonoscopy described in this book for the first time combined with MTLC may turn out to be to drug industry what nuclear reactor engineering has been to nuclear power industry (see the last row in Table 18.2). Based on this comparison, it may be asserted that:

Trying to develop safe and efficacious drugs without a comprehensive molecular theory of the living cell is akin to trying to construct a safe and efficient nuclear power plant without quantum mechanics.

(18.3)
<table>
<thead>
<tr>
<th>Final product</th>
<th>Electricity</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>By products</td>
<td>Harmful radiation and heat</td>
<td>Thousands of chemicals synthesized and discarded for every drug reaching the market; besides, only 50% of the approved drugs work</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Object</th>
<th>Plant Cell Structure</th>
</tr>
</thead>
</table>

The emergence of personalized medicine over the past decade or so has been strongly motivated by the revolutionary developments in the so-called omics in basic biological sciences throughout the second half of the twentieth century and the first decade of the twenty-first century. The development of personalized medicine is now officially endorsed by the Obama Administration. Early in 2010, Dr. M. Hamburg, FDA Commissioner, “announced a new partnership with the National Institutes of Health that is designed to more quickly get scientific and medical breakthroughs in personalized medicine, among other fields, into clinical practice” (Personalized Medicine Coalition Newsletter, Spring, 2010, p. 8).

Omics is defined as the genome-wide studies of genes (genomics), transcripts (transcriptomics), proteins (proteomics), and metabolites (metabolomics). Transcriptomics, the study of genome-wide alterations of RNA levels inside the cell, is often omitted in defining personalized medicine, but I predict that transcriptomics will play a major role in drug discovery research and personalized medicine because RNA molecules can serve as convenient intracellular reporter molecules whose behaviors can be easily monitored using ribonoscopy (Sect. 12.8.2). Another component not included in the current definition of personalized medicine is intracellular (ic) dissipative structures or ic-dissipatons (Sect. 12.5) such as the patterns of the changes in RNA levels and RNA sequences that can be used to differentiate subtypes of cancer cells (for a recent review, see van’t Veer and Bernards 2008). As can be seen in the Bhopalator models of the living cell (Fig. 2.11) and biological evolution (Fig. 14.7), all intracellular processes involving genes, RNAs, proteins, and metabolic pathways eventually converge to generate various ic-dissipatons, which are postulated to be synonymous with cell functions (see Sect. 10.2). For convenience, the study of genome-wide ic-dissipatons may, therefore, be referred to as ic-dissipatonic, in analogy to “electronics,” the study of electrons. The molecular theory of the living cell developed in this book can provide the theoretical framework for developing personalized medicine as defined in Fig. 18.3.

The other side of the personalized-health coin is personalized pathology, and according to Fig. 18.3, personalized pathology can arise in five distinct ways – due to the failure of any one of the cellular components belonging to genomics (G), transcriptomics (T), proteomics (P), metabolomics (M), or ic-dissipatomics (D), whichever happens to be the weakest link, W, in the complex network of molecular interactions constituting the living cell under a particular environmental condition. If we express the stability of the weakest ic-dissipatons in diseased cells as P(D) (the more stable the ic-dissipatons, the higher would be the probability of finding that dissipations), where P(D) is determined by the probability of the weakest G, T, P, or M, depending on the health condition of the patient, we can express the death rate, d, of individuals from the disease under consideration as in Eq. 18.4 in analogy to the death-rate equation discussed in the context of the MTLC-based model of evolution, Eq. 14.40:

\[ d = d_0[1 - P(D)] \]  (18.4)
where $d_0$ is the death rate without any therapy. In other words, diseased cells treated with drugs will have higher $P(D)$ values than diseased cells without drug treatment and hence will have lower death rate. It should be pointed out that D in Eq. 18.4 is synonymous with Node E in Fig. 14.7. Furthermore, replacing $P(D)$ with $P_{nat}^{(e)}$ in Eq. 18.1 leads to the Zeldovich-Shakhnovich model of evolution, Eq. 14.32. $P_{nat}^{(e)}$ is the probability of the native conformation of the protein that serves as the weakest link under the $i^{th}$ environmental condition (Sect. 14.7). It is quite surprising and unexpected to find that the basic mathematical equation developed to account for biological evolution from the molecular perspective, namely, the Zeldovich-Shakhnovich model, Eq. 14.32, can be applied to describe the effect of drug therapy on the death rate of human patients.