Journal of Neuroscience Methods 224 (2014) 88–95

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

A method to identify tissue cell subpopulations with distinct multi-molecular profiles from data on co-localization of two markers at a time: the case of sensory ganglia

Lugi Catacuzzeno[a,*], Luigi Sforna[a], Maria Cristina D'Adamo[b], Mauro Pessia[b], Fabio Franciolini[a]

[a] Dipartimento di Chimica, Biologia e Biotecnologie, Universita' di Perugia, via Pascoli 1, Perugia, Italy
[b] Dipartimento di Medicina Sperimentale, Sezione di Fisiologia e Biochimica, Universita' di Perugia, Perugia, Italy

HIGHLIGHTS

- It is often important to subgroup tissue cells with distinct multi-molecular profiles.
- Unfortunately, this is complicated by limitations of existing experimental techniques.
- We set an analysis to estimate the frequency of multi-molecular profiles.
- We successfully applied the analytic method to data from adult rat sensory neurons.
- We developed and make available a simple software that apply the proposed analysis.

ARTICLE INFO

Article history:
Received 3 July 2013
Received in revised form
23 November 2013
Accepted 31 December 2013

Keywords:
Molecular profile
Tissue cell subpopulations
Immunohistochemistry
Immunocytochemistry
In situ hybridization
Sensory neurons

ABSTRACT

Background: Most biological tissues are characterized by high morphological and functional cell heterogeneity. To investigate this heterogeneity at the molecular level, scientists have tried to associate specific sets of molecular markers (molecular profiles) to functionally distinct cell subpopulations, evaluating their expression using immunochemistry and in situ hybridization techniques.
New method: We propose here a novel analysis that allows the estimation of the frequency of cells expressing distinct molecular profiles starting from data on the co-expression of two markers at a time. In order to facilitate the application of the proposed analysis, we developed and make available a user-friendly window-based software.
Results: We successfully applied the analytical method to experimental data from adult rat sensory neurons. In a first application we subgrouped DRG neurons in 11 subpopulations on the basis of the co-expression of 6 molecular markers (the TRPs type V1, A1, and M8 and the trks type A, B, and C). In a second application we found that while rat DRG have significant frequencies of peptidergic/IB4-negative and non-peptidergic/IB4-positive nociceptors, rat TG neurons lack almost completely these two subpopulations.
Comparison with existing methods: The analytical method here proposed overcomes the limitations of the presently available experimental techniques, most of which can assess the co-expression of only few molecular markers at a time.
Conclusions: This new method will allow a better understanding of the molecular and cellular heterogeneity of tissues in normal and pathological conditions.

1. Introduction

Most biological tissues are characterized by high morphological and functional cell heterogeneity. This heterogeneity must be investigated if one wishes to understand and correlate the phenotypes of the cell subpopulations present in a given tissue with their biological functions. A typical case is given by sensory ganglia which display more than twenty types of functionally distinct sensory neurons, each transducing a specific form of stimulating energy and

Abbreviations: DRG, dorsal root ganglion; TG, trigeminal ganglion; TRP, transient receptor potential; IB4, isolectin B4; trk, tyrosine receptor kinase; LPPS, linear programming problem solver; CGRP, calcitonin gene receptor peptide; SP, substance P.

* Corresponding author at: Dipartimento di Chimica, Biologia e Biotecnologie, Universita' di Perugia, via Pascoli 1, Perugia, Italy. Tel.: +39 75 585 5750.
E-mail address: luigi.catacuzzeno@unipg.it (L. Catacuzzeno).

0165-0270/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.12.015

transporting the resulting signal (i.e. a train of action potentials) along the axon with a specific conduction velocity (Craig, 2003; Ma, 2010; Perl, 2007; Darian-Smith, 1984).

In the attempt to understand the function of the various subpopulations of sensory neurons at the molecular level, scientists have looked for proteins relevant in the transduction and conduction of the various sensory modalities, including ion channels, neuropeptides, transmitters, calcium-binding proteins, and neurotrophic receptors. It was found, for instance, that ion channels belonging to the transient receptor potential type M and V are gated by temperature changes, and are candidate transducers of skin temperature in both the noxious and non-noxious ranges (Lumpkin and Caterina, 2007; Julius and Basbaum, 2001; Reid, 2005). The neurotransmitters substance P (SP) and calcitonin gene-related peptide (CGRP) have instead been linked to nociceptive function, as both were expressed in a higher proportion of nociceptive than non-nociceptive somatic afferent neurons (Lawson, 2002).

These and other studies clearly show that functionally relevant molecular markers are differentially expressed in specific subgroups of sensory neurons, in accordance with the notion that the functional heterogeneity of sensory neuron subpopulations originates from the heterogeneity in the set of proteins expressed. However, the same studies also show that no single marker neatly associates with a specific neuronal type, suggesting that the functional properties of a specific cell subpopulation depend on the particular combination of the markers expressed (Belmonte and Viana, 2008; Gold et al., 1996). To define the degree of co-expression of several different molecular markers (i.e., to define the molecular profile) for each sensory neuron subpopulation is not easy. Even the assessment of the full set of molecular profiles resulting from the use of four markers, although possible in principle, is at the moment experimentally prohibitive, as most published data limit the assessment of markers co-expression to few (two, at most three) at a time.

In the attempt to overcome these limitations and define the molecular profiles of the various cell subpopulations present in biological tissues, here we propose a new analysis capable to assess the frequency of occurrence of all the possible molecular profiles, using as input data those obtained from experiments assessing the co-expression of two markers at a time. To facilitate the application of the reported analysis, we developed a simple, ready-to-use, window-based software, available as supplementary data of this paper.

2. Theory

Let us consider N different markers that can be either expressed or not expressed in the cells composing the tissue under study. In terms of the considered markers, there are 2^N possible molecular profiles that cells can belong to. As a practical example, if we consider three markers A, B, and C, there will be 8 (i.e. 2^3) possible molecular profiles, namely $A^+B^+C^+$, $A^+B^+C^-$, $A^+B^-C^+$, $A^+B^-C^-$, $A^-B^+C^+$, $A^-B^+C^-$, $A^-B^-C^+$, and $A^-B^-C^-$. Our goal is to estimate the fraction of cells within the population under study that is characterized by each of the possible molecular profiles, i.e. to determine the frequencies f_i = number of cells showing the molecular profile i/total number of cells, where i represents one of the possible 2^N profiles, $\sum_i f_i = 1$, and $0 \le f_i \le 1$ for all i. In our problem there are two types of known experimental (input) data: (1) the frequencies of cells expressing each of the N markers, P_i, where i represents one of the considered markers; (2) the frequencies of cells co-expressing two given markers i and j, P_{ij}, where $i \neq j$. Thus our problem consists of 2^N unknowns (the f_i) and at most $N + (N!/2(N-2)!) + 1$ input data, i.e. the N P_i's, the $N!/2(N-2)!$ P_{ij}'s and the additional condition that the sum of all profile frequencies must be unity. The number of input data will be lower than $N + (N!/2(N-2)!) + 1$ when not all possible co-expressions between two markers are experimentally tested.

To better understand the properties of our problem, we can express it as a system of linear equations, that is conveniently written down in the following matrix form:

$$Ax = b$$

In the above equation x represents a vector of size 2^N containing the unknown frequencies f_i, b represents a vector of size $M \le N + (N!/2(N-2)!) + 1$ containing the input data. Finally, A is a $N \times M$ coefficient matrix, having all terms equal to either 0 or 1. For the case of the three markers A, B, and C reported above the matrix form of our problem will be

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_{A^+B^+C^+} \\ f_{A^+B^+C^-} \\ f_{A^+B^-C^+} \\ f_{A^+B^-C^-} \\ f_{A^-B^+C^+} \\ f_{A^-B^+C^-} \\ f_{A^-B^-C^+} \\ f_{A^-B^-C^-} \end{bmatrix} = \begin{bmatrix} P_A \\ P_B \\ P_C \\ P_{AB} \\ P_{AC} \\ P_{BC} \\ 1 \end{bmatrix}$$

For the case of interest here, where $M < N$, the system of linear equations is underdetermined, and can either have no solutions (in the case that two or more equations are incompatible) or infinite solutions. In the latter case the solution space can be written as a particular solution x_p added to any linear combination of N–M vectors v_i, which are said to be in the null-space of matrix A

$$x = x_p + s_1 v_1 + s_2 v_2 + \cdots + s_{N-M} v_{N-M}$$

In the above equation the $s_1, s_2, \ldots, s_{N-M}$ can vary from $-\infty$ to $+\infty$, thus allowing each x_i to assume all real values. However in our particular problem not all solutions x have a realistic meaning, since for many of them the frequency associated to one or more molecular profiles will unrealistically assume values outside the range $0 \le f_i \le 1$. Consideration of the above reported condition for all f_i will result in a restriction of the range of possible solutions x. Our problem is thus to find the extremes of the allowed ranges for each f_i, by solving the linear system of equations $Ax = b$ subject to the conditions $0 \le x \le 1$. This problem can be conveniently expressed as the following linear programming problem:

For all $i = 1, 2, \ldots, 2^N$
Minimize: $x_i^{\min} = x_i$
Subject to:
$Ax = b$
and $0 \le x \le 1$
And maximize: $x_i^{\max} = x_i$
Subject to:
$Ax = b$
and $0 \le x \le 1$

for which efficient computational algorithms exist (Press et al., 1992). The above linear programming problem is however still incomplete for our purpose, since we do not exactly know the input data b needed to find a solution. We have indeed only an experimental estimate of b, b^*, that differs from b because of experimental errors, err. In other words, we only know that the vector b will lie with high probability somewhere between $b^* - err$ and $b^* + err$. A meaningful value for err could for example be given by the experimentally estimated standard errors for b. By including