Models of cell signalling uncover molecular mechanisms of high-risk neuroblastoma and predict outcome

Marta R. Hidalgo[1], Alicia Amadoz[2], Cankut Çubuk[1], José Carbonell-Caballero[2] and Joaquín Dopazo[1,3,4]

1 Clinical Bioinformatics Research Area. Fundación Progreso y Salud (FPS). Hospital Virgen del Rocio. 41013. Sevilla. Spain

2 Computational Genomics Department. Centro de Investigación Principe Felipe (CIPF). 46012 Valencia. Spain

3 Functional Genomics Node (INB). FPS. Hospital Virgen del Rocio. 41013 Sevilla. Spain

4 Bioinformatics in Rare Diseases (BiER). Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER). FPS. Hospital Virgen del Rocio. 41013. Sevilla. Spain;

Abstract

Despite the progress in neuroblastoma therapies the mortality of high-risk patients is still high (40% - 50%) and the molecular basis of the disease remains poorly known. Here we use models of cell signalling, a key process in this cancer, to understand the molecular determinants of bad prognostic. We also show how the activity of signalling circuits can be used as a predictor of survival in neuroblastoma patients.

Introduction

Neuroblastoma is a tumour derived from primitive cells of the sympathetic nervous system that, despite advances in its treatment still has a poor survival for high-risk patients (Brodeur, 2003). Risk groups are defined according to disease stage, patient age, and *MYCN* amplification status (Øra and Eggert, 2011). Although the use of biomarkers has demonstrated a clear clinical utility, they represent statistical associations to clinical parameters and frequently lack any mechanistic relationship with the molecular mechanisms responsible for tumorigenesis or therapeutic response. On the contrary, signalling pathways play a key role in these processes. Actually, it has recently been demonstrated in neuroblastoma (Fey, et al., 2015) and other cancers (Hidalgo, et al., 2017) that the activity of specific circuits of signalling pathways was more correlated to patient survival that any of their constituent genes.

Here, we have used models that produce a realistic estimation of signalling circuit activity within pathways from gene expression (Hidalgo, et al., 2017) to discover the molecular mechanisms behind the differences between the patients with *MYCN* amplification and the others as well as the determinants of survival in neuroblastoma.

Results

Data were downloaded and processed as indicated in Methods. Overall, our results document extensive differences at the level of signalling activity between patients with *MYCN* amplification and those lacking this biomarker. Specifically, the MAPK and NFKβ pathways present a dramatic reduction of activity in patients with *MYCN* amplification. Some relevant functions significantly deactivated in high-risk patients are, for example, *cellular response to DNA damage stimulus* (see Figure 1), *negative regulation of cell division* or *cell cycle arrest*.

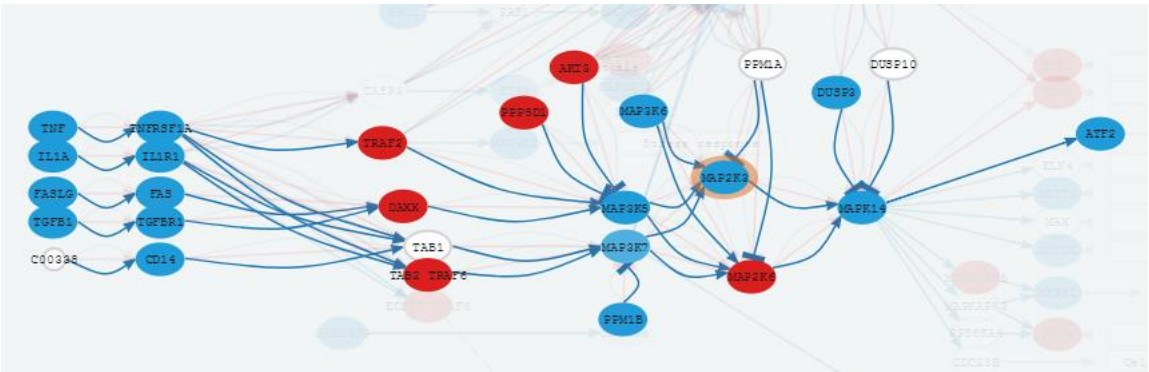

Figure 1. *Circuit MAPK signalling pathway: ATF2 that triggers cellular response to DNA damage stimulus is significantly downregulated (FDR-adjusted p-value=1.45x10^{-15}*

Across different pathways high-risk patients present a systematic deactivation of cellular functions related to DNA repair and immune response.

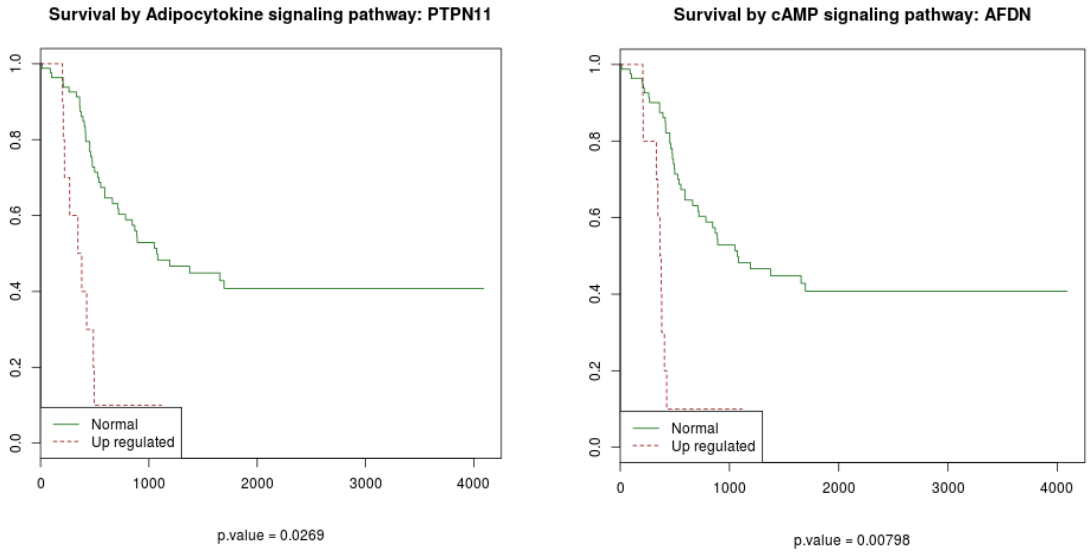

Figure 2*. Kaplan-Meyer plots of survival of patients with MYCN amplification which have downregulated Adipocytokine: PTPN11 (left) and cAMP: AFDN (right) signalling circuits*

Regarding survival across all the patients, processes such as *negative regulation of apoptosis process*, triggered by multiple signalling circuits: *p53: CDK1 CCNB3* (FDR-

adjusted p-value=7.6x10^{-11}) *PI3K-Akt*: *BCL2L1** (adj. p-val.=5.1x10^{-7}) *Jak-STAT*: *RAF1* (adj. p-val.=5.4x10^{-3}) *Fc epsilon RI: MAPK8* (adj. p-val.=5.5x10^{-3}), is associated to bad prognostic. Similarly, *negative regulation of cell cycle arrest* (triggered by *p53*: *MDM2**, adj. p-val. <10^{-20}) and positive regulation of cell growth (*Sphingolipid*:*TP53 BCL2,* adj. p-val.= 1.5x10^{-12}) seem to play an important role in low survival.

We also tried to find the molecular drivers of bad prognostic among patients with *MYCN* amplifications. Only two circuits, *Adipocytokine: PTPN11* and *cAMP: AFDN* are clearly associated to bad prognostic. One of the effector proteins, *PTPN11* has been implicated in mitogenic activation, metabolic control, transcription regulation, and cell migration (Chan, et al., 2008). The other effector protein, *AFDN*, is the fusion partner of acute lymphoblastic leukemia (*ALL-1*) gen involved in acute myeloid leukemias with t(6;11)(q27;q23) translocation, with a known role in cell adhesion (Mandai, et al., 2013).

Conclusions

The use of models that quantify cell behavioural outcomes provides a unique opportunity to understand the molecular mechanisms of cancer development and progression as well as open the possibility to highly specific, individualized therapeutic interventions.

Methods

Data Source and data preprocessing

The matrix GSE49711_SEQC_NB_TUC_G_log2.txt, with gene expression levels estimated by Cufflinks (Trapnell, et al., 2012) and quantified as $\log_2$(1+FPKM), was downloaded from the GEO database. Batch effect was corrected with COMBAT (Johnson, et al., 2007). Finally, the values were normalized between 0 and 1.

Signalling circuit activity model

Circuit activities are modelled from gene expression values as described in (Hidalgo, et al., 2017). Briefly, KEGG pathways (Kanehisa, et al., 2014) are used to define circuits connecting receptor proteins to effector proteins. A total of 98 KEGG pathways involving a total of 3057 gene products that compose 4726 nodes were used to define a total of 1287 signalling circuits. Normalized gene expression values are used as proxies of protein activity (Efroni, et al., 2007; Montaner, et al., 2009; Sebastian-Leon, et al., 2014). The signal transmission is estimated by starting with an initial signal of 1, which is propagated along the nodes of the signalling circuits according to the following recursive rule:

$$S_n = v_n \cdot \left(1 - \prod_{s_a \in A}(1 - s_a)\right) \cdot \prod_{s_i \in I}(1 - s_i) \quad (1)$$

Where S_n is the signal intensity for the current node n, v_n is its normalized gene expression value, A is the total number of activation signals (s_a), arriving to the current node from activation edges, I is the total number of inhibitory signals (s_i) arriving to the node from inhibition edges (Hidalgo, et al., 2016). In addition to circuit activities, the signal received by specific cell functions -according to either Gene Ontology (Ashburner, et al., 2000) or Uniprot (UniProt_Consortium, 2015) definitions-, triggered by more than one circuit, can also be estimated.

Survival analysis

Kaplan-Meier (K-M) curves (Kaplan and Meier, 1958) are used to relate module activity to patient survival in the different cancers. The value of the activity estimated for each module in each individual was used to assess its relationship with individual patient survival. Calculations were carried out using the function *survdiff* from the *survival* R package (https://cran.r-project.org/web/packages/survival/).

References

Ashburner, M.*, et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet*, **25**, 25-29.

Brodeur, G.M. (2003) Neuroblastoma: biological insights into a clinical enigma, *Nature Reviews Cancer*, **3**, 203-216.

Chan, G., Kalaitzidis, D. and Neel, B.G. (2008) The tyrosine phosphatase Shp2 (PTPN11) in cancer, *Cancer and metastasis reviews*, **27**, 179-192.

Efroni, S., Schaefer, C.F. and Buetow, K.H. (2007) Identification of key processes underlying cancer phenotypes using biologic pathway analysis, *PLoS ONE*, **2**, e425.

Fey, D.*, et al.* (2015) Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients, *Sci Signal*, **8**, ra130.

Hidalgo, M.*, et al.* (2016) High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes, *Oncotarget*, **In press**.

Hidalgo, M.R.*, et al.* (2017) High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes, *Oncotarget*, **8**, 5160-5178.

Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods, *Biostatistics*, **8**, 118-127.

Kanehisa, M.*, et al.* (2014) Data, information, knowledge and principle: back to metabolism in KEGG, *Nucleic Acids Res*, **42**, D199-205.

Kaplan, E. and Meier, P. (1958) Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457-481.

Mandai, K.*, et al.* (2013) Afadin/AF-6 and canoe: roles in cell adhesion and beyond, *Progress in molecular biology and translational science*, **116**, 433-454.

Montaner, D.*, et al.* (2009) Gene set internal coherence in the context of functional profiling, *BMC Genomics*, **10**, 197.

Øra, I. and Eggert, A. (2011) Progress in treatment and risk stratification of neuroblastoma: impact on future clinical and basic research. *Seminars in cancer biology*. Elsevier, pp. 217-228.
Sebastian-Leon, P.*, et al.* (2014) Understanding disease mechanisms with models of signaling pathway activities, *BMC Syst Biol*, **8**, 121.
Trapnell, C.*, et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, *Nat Protoc*, **7**, 562-578.
UniProt_Consortium (2015) UniProt: a hub for protein information, *Nucleic Acids Res*, **43**, D204-212.