Universidad de Sevilla

Facultad de Biología

Departamento Biología celular

Topological analysis of the tumour microenvironment to study Neuroblastoma

Tesis doctoral en el programa de Biología integrada

Pablo Vicente Munuera

Sevilla, 2020

Doctorado

Pablo Vicente Munuera

Director/Tutor

Luis M. Escudero Cuadrado

Agradecimientos

A Mónica, gracias.

A Luisma, gracias.

A mi familia, gracias.

A Pedro, gracias.

A Antonio, gracias.

A Carmen, gracias.

A Ana, gracias.

A Rebeca Burgos Panadero, gracias.

Al grupo de Alberto Pascual y a él mismo, gracias.

A Beatriz Estrada, gracias.

A Valentina Annese, gracias.

A los frikis informáticos, gracias.

A los comensales de Biología Celular y la charla de cada día: "My husband won't fit" es la mejor serie ever, gracias.

A Reme, gracias.

A Ángel Velasco, gracias.

A toda la gente que ha pasado por el laboratorio: Charlie, Tina, Antonio V., Victor, Ana, Irene, Jesús, Veredas, Pedro H., ...

A Francisco Escolano, gracias.

A Ernesto Estrada, gracias.

A todas y cada una de las demás personas que, para bien o para mal, han estado alguna vez en mi vida: gracias, habéis hecho que sea quien soy ahora.

Summary

Solid tumours and their tumour microenvironment (TME) can be considered as complex networks whose elements are in constant physical stress. All the elements of the TME, including tumour cells, stromal cells, immune and stem cells, blood/lymphatic vessels, nerve fibers and extracellular matrix components, belong to a highly balanced compression-tension molecular and cellular structure. Through mechanical signals, each element could affect its surroundings modulating tumour growth and migration. The analysis of these complex interactions and the understanding of the structural organization of a tumour requires the collaboration of different disciplines. In this thesis, we focus on a particular solid tumour: Neuroblastoma, a rare type of cancer, originated during the embryo development. We apply computational and mathematical tools to analyse the topology of vitronectin, a glycoprotein of the extracellular matrix, in neuroblastoma tumours. Vitronectin has a particular interest in tumour biology where it is associated with cell migration, angiogenesis, and matrix degradation. Still, its role in Neuroblastoma is not clear. Here, we study the organization of vitronectin within the TME considering Neuroblastoma patient prognosis and tumoral aggressiveness. Combing graph theory and image analysis, we characterize histopathological images taken, from a human sample, by analysing different topological features that capture the organizational cues of vitronectin. By means of statistical analyses, we find that two topological features (Euler number and branching), related to the organization of the existing vitronectin within and surrounding the cells (territorial), correlates with risk pre-stratification group and genetic instability criterion. We interpret that a large amount of recently synthesized VN would create tracks to aid malignant neuroblasts to invade other organs, pinpointed by both topological features, which in turn would change, dramatically, the constitution and mechanics of the extracellular matrix, increasing tumour aggressiveness and worsen patient outcomes. Further studies will be required to assess the true potential of vitronectin as a future therapeutic target of neuroblastoma.

Resumen

Los tumores sólidos y su microambiente tumoral (TME) pueden ser vistos como redes complejas cuyos elementos están en constante estrés físico. Todos los elementos del TME, incluidas células tumorales, células del estroma, células inmunes y células troncales, vasos sanguíneos o linfáticos, fibras nerviosas y componentes de la matriz extracelular, pertenecen a una maquinaria molecular y celular de tensión-compresión altamente equilibrada. A través de señales mecánicas, cada elemento podría afectar su entorno modulando el crecimiento tumoral y la migración. El análisis de estas interacciones complejas y la comprensión de la organización estructural de un tumor requiere la colaboración de diferentes disciplinas. En esta tesis, nos centramos en un tumor sólido particular: el neuroblastoma, un cáncer considerado como 'raro', que se origina durante el desarrollo del embrión. Aplicando herramientas computacionales y matemáticas, analizamos la topología de la vitronectina, una glicoproteína de la matriz extracelular, en tumores de neuroblastoma. La vitronectina tiene un interés particular en la biología tumoral, ya que está asociada con migración celular, angiogénesis y degradación de la propia matriz. Aún así, su papel en el neuroblastoma no está claro. En este trabajo, estudiamos la organización de la vitronectina dentro del microambiente tumoral, considerando el pronóstico del paciente con neuroblastoma y su agresividad tumoral. Combinando la teoría de gráficos y el análisis de imagen, caracterizamos las imágenes histopatológicas tomadas de una muestra humana, mediante el análisis de diferentes características topológicas que capturan la organización de la vitronectina. Mediante análisis estadísticos, encontramos que dos características topológicas (número de Euler y 'ramificación'), relacionadas con la organización de la vitronectina existente dentro y alrededor de las células (territorial), se correlacionan con el grupo de pre-estratificación de riesgo y la inestabilidad genética del paciente. En consecuencia, interpretamos que una gran cantidad de VN, sintetizada recientemente, crearía una especia de 'caminos' para ayudar a los neuroblastos malignos a invadir otros órganos, que a su vez cambiarían dramáticamente la constitución y la mecánica de la matriz extracelular, aumentando la agresividad del tumor y empeorando el pronóstico del paciente. Futuros estudios serán requeridos para evaluar el verdadero potencial de la vitronectina como una diana terapéutica del neuroblastoma a largo plazo.

INDEX

Abbreviations

AIC	Akaike information criterion.
Chi-sq	chi-square (χ^2).
ECM	extracellular matrix.
GAGs	glycosaminoglycans.
GDD	graphlets degree distribution.
INRG	International Neuroblastoma Risk Group.
MNA	MYCN amplified.
MST	minimum spanning tree.
NB	neuroblastoma.
ORCA	orbit counting algorithm.
PAI-1	plasminogen activator inhibitor-1.
SCAs	segmental chromosome aberrations.
STD	standard deviation.
TME	tumour microenvironment.
uPAR	urokinase plasminogen activator receptor.
VN	vitronectin.

Definitions

Tensegrity: Stabilization of structures constituted by continuous elements of tension and discontinuous elements resistant to compression.

Topology: How a set of elements are structured and connected in a given space.

Tessellation: A surface covered by geometric components (or tiles) with no gaps and without overlapping.

Graph/network: A set of elements connected between them following determined rules that represent binary relations. A graph is formed by nodes (the elements) and edges that link them.

Voronoi diagram: A particular tessellation formed by convex polygons. Each convex polygon is a Voronoi cell. Every Voronoi cell emerges from a seed. All the points of a Voronoi cell are closer to its own seed that to any other seed of the surface.

Graphlets: Graphs with a small number of nodes extracted from a larger network. A network can be quantitively characterized by its graphlets composition.

Mark-up image: A binary immunohistochemistry microscopic image in which the white regions represent detected objects, and the black ones, the background.

Euler number feature: In an image where a set of objects has been identified, the Euler number is the value of the number of objects minus the number of holes inside them.

Branching feature: In an image where a set of objects has been identified, the Branching is the value of the number of crosslinks that are found on the objects.

Node: A representation of an object. In our case, it stands for a hexagonal area filled with vitronectin.

Edge: The link between nodes. Two nodes connected by one edge are considered adjacent.

Tensegrity index: Represents how different is the VN (interterritorial or territorial) organized in the biopsy compared with a homogenous distribution.

1 Introduction

Evolution gifted animals with eyes to study and interact with their environment. We, humans, have developed a vision capable of detecting highly complex situations. It was a matter of time before we discovered how to stop time by painting or, later, capturing images using photography. We, then, invented the microscopy to scrutinise our body in more detail, letting us see tiny portions of our body like the cells. In the 20th century, the only way to analyse these situations was by an expert's eyes. However, we now can obtain more precise information by acquiring a larger quantity of images with higher resolution. In the same way, as many experimental fields grow, such as biology and medicine, they are generating a huge amount of information that needs to be thoroughly quantified. In any case, a human no longer can process that overwhelming number of images in a timely manner.

Nowadays, collaboration between different sciences is more important than ever to keep science moving toward progress. Scientific fields that may appear extremely far away are now doing innovative science together. This can also be applied to the fields of biology and medicine, where, for instance, physical and mathematical works are common (Fiorino et al., 2014; Gómez-Gálvez et al., 2018; Tsuboi et al., 2018). In these studies, teamwork and different point of views need to be balanced, as in a tissue happens, where many components are at work to achieve homeostasis. A fitting example is the expanding field of computational biology and cancer, where computerized image analysis is now improving and increasing its methods as a result of the growing demand for objectively quantify raw data (Guirao et al., 2015; Heller et al., 2016; Vicente-Munuera et al., 2020).

Computerized image analysis exploits several aspects of an image, which can be englobed into three categories depending on the features they extract: texture properties, morphological characteristics and graph features (topology) (Belsare and Mushrif, 2012). Texture analysis is based on the intensity and colours of the image. In particular, texture features capture the periodicity and scale of some regional pattern by means of direction, coarseness or contrast (Howarth and Rüger, 2004). A successful example of texture analysis was a comparison of tomography images from primary lung tumours and granulomatous lesions (Dennie et al., 2016). Another category is morphology, which considers the shape of the detected elements in the image. For instance, a morphological feature can capture the size, radius, perimeter, among other parameters, of the objects of

an image. Some of these features turned out to be relevant to assess the aggressiveness of neuroblastic tumours capturing blood vessels patterns (Tadeo et al., 2018). Last but not least, graph-based characteristics (or topology) are related to how the elements of an image are particularly structured and connected. For example, in (Sanchez-Gutierrez et al., 2013) identified graph-related attributes that characterized the epithelial organization and rearrangements during development. Overall, computer tools allow a rapid and effective analysis of biological samples, reducing inter and intra-observer variability and placing medicine and biology in the way of resolving the most complex puzzles (Rojo et al., 2010).

1.1 NETWORKS

Historically, networks theory (or graph theory) has been successfully applied to social sciences (Scott, 1988) and computational sciences (Albert et al., 1999; Holme et al., 2002). In the past years, this tendency has shifted to biological environments thanks to multidisciplinary teams, where a plethora of studies using network theory to analyse biological systems was published (Barabási and Oltvai, 2004; Binchi et al., 2014; Escudero et al., 2011; Sáez et al., 2013). In particular, protein-protein interaction has been a fruitful niche due to the development of high-throughput techniques, such as Yeast Two-Hybrid (Escolano et al., 2012; Estrada, 2006; Fields and Song, 1989; Jeong et al., 2001; Pržulj et al., 2004). Regarding biological images, graph theory is used to model how the specified elements are organized throughout a sample. Voronoi diagrams, Delaunay triangulations and minimum spanning tree algorithms have been recurrently used to construct a network from elements like the cell nuclei (Angel Arul Jothi and Mary Anita Rajam, 2017; Doyle et al., 2008; Yao et al., 2015). In other cases, like epithelial tissues, it is relatively easy to build a network by connecting the cells that are neighbours (Escudero et al., 2011; Sanchez-Gutierrez et al., 2013).

We could provide many definitions of a network (or graph) (Boccaletti et al., 2006; Chung and Graham, 1997; Kay et al., 1977; Strogatz, 2001). However, a network is nothing, but several elements connected in a pairwise manner (**Figure 1**). As in social life, in a graph, all lies on connections and how, we, the nodes (**Figure 1,** circles), are suitably linked by edges (**Figure 1,** connection between circles). In the case of a relevant person, many people are, usually, in contact with him/her, transforming him/her into a highly linked node and, therefore, a hub. Hubs turned out to be important in biology because many paths pass through them (He and Zhang, 2006; McCormack et al., 2016). Depending on the structure a network displays, they can be categorized into a random network, small-world or scale-free, among others (Newman, 2003;

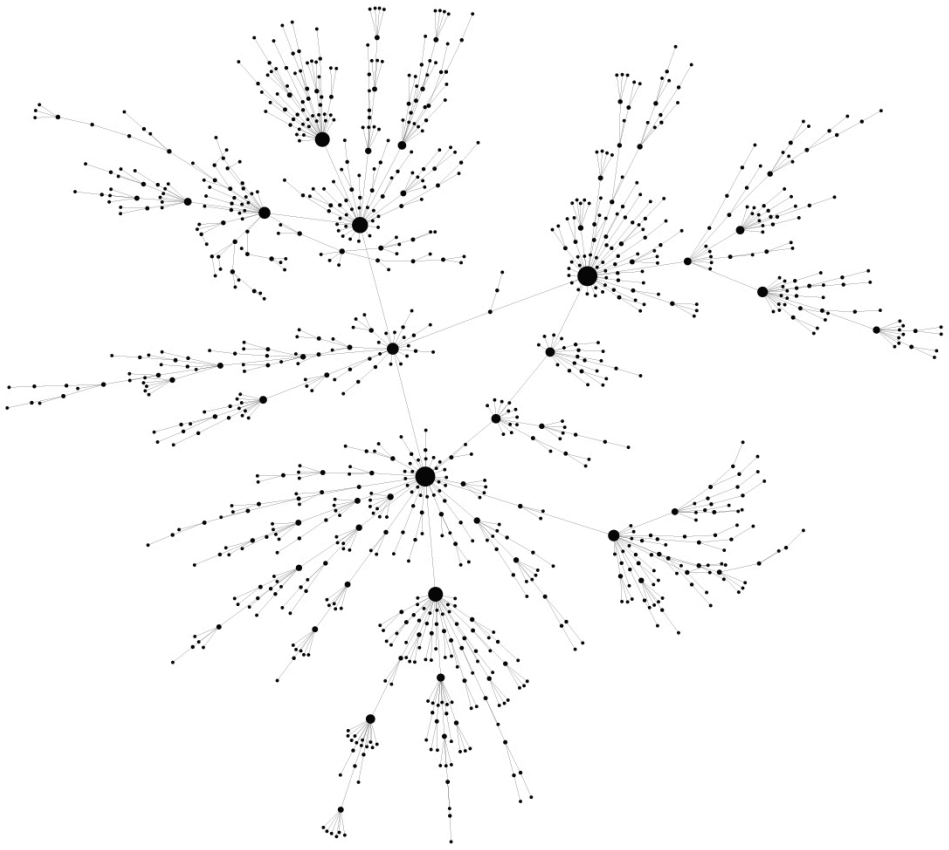

Figure 1. **Example of an scale-free random network**. It was generated from Barabasi-Albert algorithm with 1000 nodes, computed with the plugin "Network randomizer" of Cytoscape (Shannon et al., 2003). The size of the nodes is related to its degree. The links (edges) between the nodes are in grey colour.

Valverde et al., 2015). Despite the wide range of structural categories, protein-protein interaction or metabolic networks were proven to have a higher tendency to display scale-free properties, where a few nodes with a higher number of edges (hubs) tend to be linked with many low-connected nodes (Barabási and Oltvai, 2004; Junker and Schreiber, 2008). Social networks, like Twitter, exhibit also scale-free properties (Ediger et al., 2010; Mislove et al., 2007).

Although a network is simply formed by nodes and its connections (edges), exists many parameters that can be used to characterize a graph (Almaas, 2007; Dong and Horvath, 2007). For example, the connectivity of a network can be analysed by means of its node degree distribution or connected components. Both features focus on how the number of edges per node are distributed, i.e. how the nodes are connected between each other. Similarly, the clustering coefficient can be used to understand the network architecture by measuring the density of local connections. Another relevant concept is network centrality. This concept

identifies the most 'relevant' node of a graph regarding properties like the degree of the nodes (degree centrality) or shortest path (betweenness centrality). For instance, considering the degree of the nodes, highly relevant nodes will be nodes with lots of connections. In a social network, famous people will be identified as 'relevant nodes', connecting many other nodes. Therefore, if we remove the node with the largest centrality, we may disconnect some nodes or, in an extreme scenario, we could obtain an unconnected network (two or more separate graphs with no links between them). The degree centrality of a network was successfully used to link network topology and protein phenotype and expression dynamics (Yu et al., 2007).

Although it is possible to compare two networks in terms of their properties, it may be more informative to compare both graph's structures. In order to do so, Pržulj and cols introduced a term called "graphlets" (**Figure 2**), which stands for small connected non-isomorphic induced subgraphs (Pržulj et al., 2004). In a later work, they also defined a formula to compute the distance between two Graphlets Degree Distribution (GDD) (Przulj, 2007) normalizing the possible differences in a range between 0 and 1. Specifically, graphlets comparisons are more focused on local structural properties rather in global ones. There are several works using graphlets as a measure (Przulj, 2007; Pržulj et al., 2004; Yaveroğlu et al., 2014) or improving

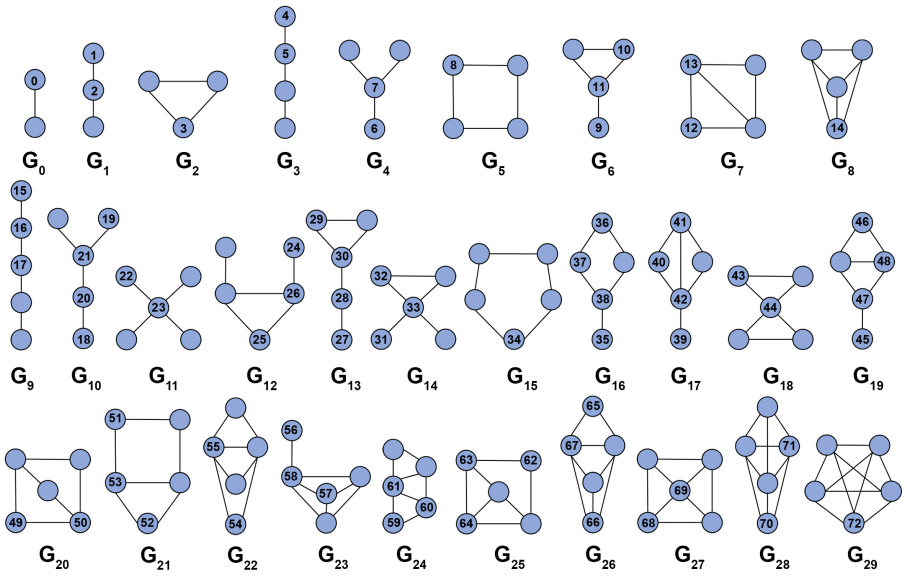

Figure 2. **Illustration of graphlets networks used in (Pržulj, 2007).** A representation of graphlets of up to five nodes. Each graphlet is labelled with a Gn, in which 'n' is the graphlets number (from G0 to G29). Inside a graphlet there are different configurations, called orbits (from 0 to 72).

them (Ahmed et al., 2017; Cannoodt et al., 2018). However, this technique has still much potential to be exploited within biology.

1.2 NEUROBLASTOMA

Neuroblastoma (NB) is a type of solid cancer, considered as a “rare disease”, that forms in certain neural tissues from the sympathetic nervous system during the development of the embryo. Nonetheless, it is the third most common tumour during infancy, after leukaemia and brain cancer (Ediger et al., 2010). NB is rare in adults: about 90% of cases occur in children younger than 5 years old (Maris, 2010). This cancer usually starts in one of the adrenal glands located above the kidneys, or in the nerve tissue located at the spinal cord in the neck, chest, tummy, or pelvis (Voûte et al., 1992). Historically, NB was first described as an undoubted tumour of the sympathetic system, designating it as a glioma by Rudolf Virchow in 1865 (Virchow, 1865) on a book called “Diseased tumours” (**Figure 3**). But in 1910, James Homer Wright finally named it Neurocytoma or Neuroblastoma, since the cells that formed the cancer were essentially undifferentiated nerve cells (Jaffe, 1976; Wright, 1910).

Thanks to scientific advances, nowadays, we have a deeper knowledge of the origin of NB tumours. Starting at the neural crest and neural tube, neuroblast progenitors begin to migrate to, eventually, diverge into a neural cell (Marshall et al., 2014). Throughout this process, the oncogene MYCN has a predominant role modulating the migration and expansion of neural cells

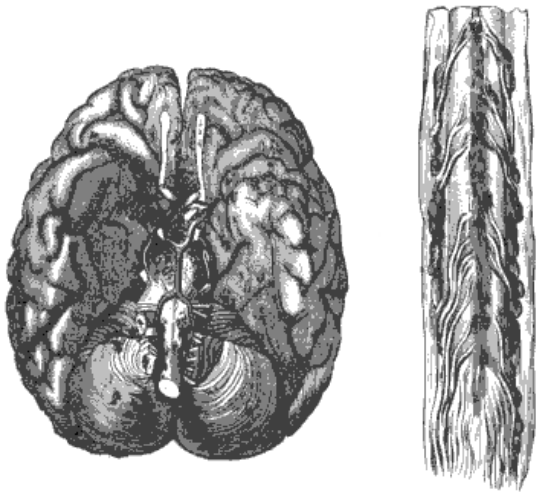

Figure 3. Illustrations extracted from Virchow’s book “Diseased tumours” of a chapter called “Melanoma of the soft meninges”. Quoting from german and translated in English: “Fig. 128 (left). Multiple melanomas of the pia mater basilaris, mostly around the medulla oblongata, the ponts, the fossa sylvii, fissura longit. Fig. 129 (right). The lower end of the spinal cord of Fig. 128 (left), with multiple melanomas of the soft skin, forming nodular swellings on the nerve roots.”

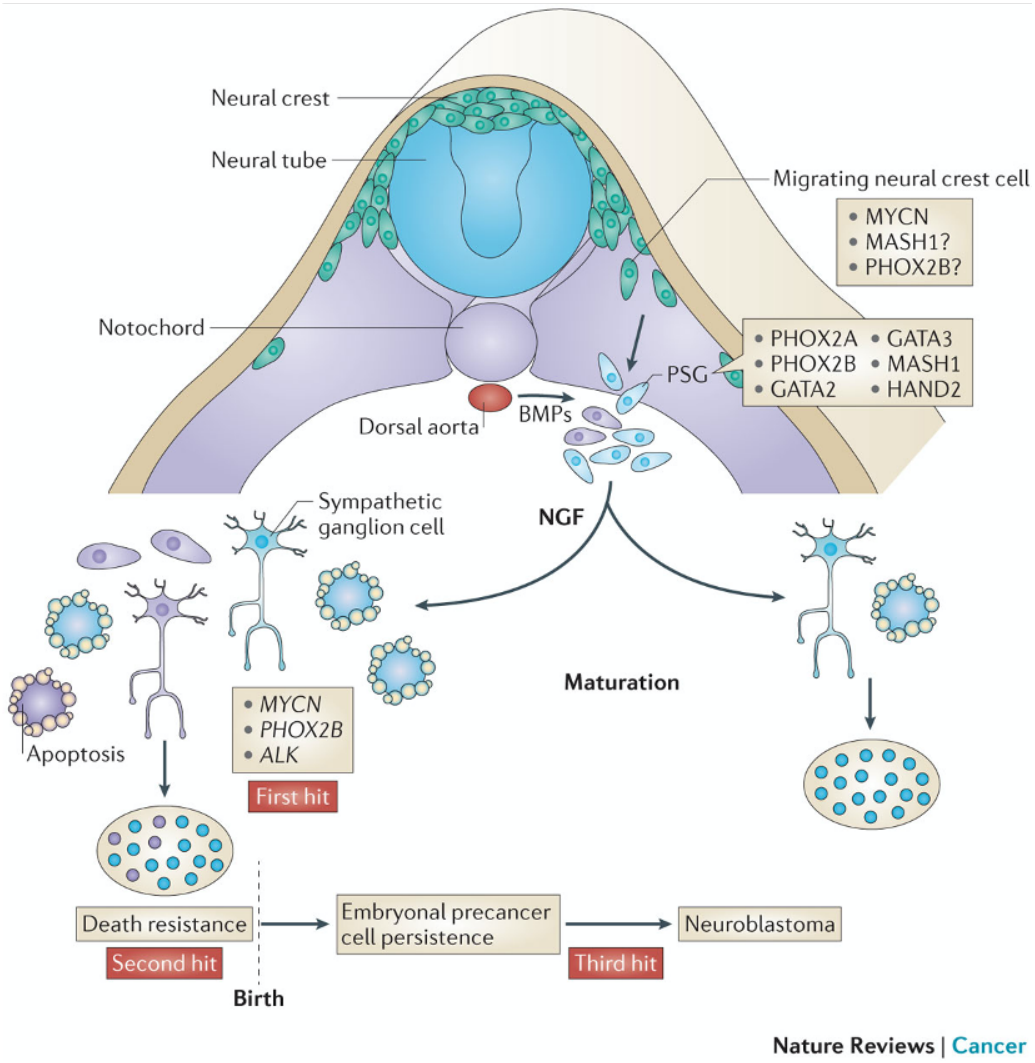

Figure 4. Schematic origin of a neuroblastoma tumour. Neuroblasts starting at the neural tube and neural crest will migrate and undergo different stages to, ultimately, become a mature cell of the neural system. Blue cells are normal cells, while purple cells are neuroblast cancer cells. PSG: Primary sympathetic ganglia; BMPs: Bone morphogenetic proteins; ALK: Anaplastic lymphoma kinase; NGF: Nerve growth factor. Taken from (Marshall et al., 2014).

(Hansford et al., 2004; Zimmerman et al., 1986). However, a large percentage of neural progenitors should undergo apoptotic cell death to maintain homeostasis (Dekkers et al., 2013). In rare instances, some cells ignore the apoptotic signals and, if they acquire additional mutations, it could lead to postnatal malignant diseases, such as NB (**Figure 4**, purple cells).

NB is a heterogeneous tumour with widely varying prognosis according to several clinical and genetic factors described in the International Neuroblastoma Risk Group (INRG) classification (**Figure 5**) (Cohn et al., 2009). Depending on the INRG category a NB patient is assigned, he/she will be treated accordingly. In particular, NB patients can be classified into very low-risk (good

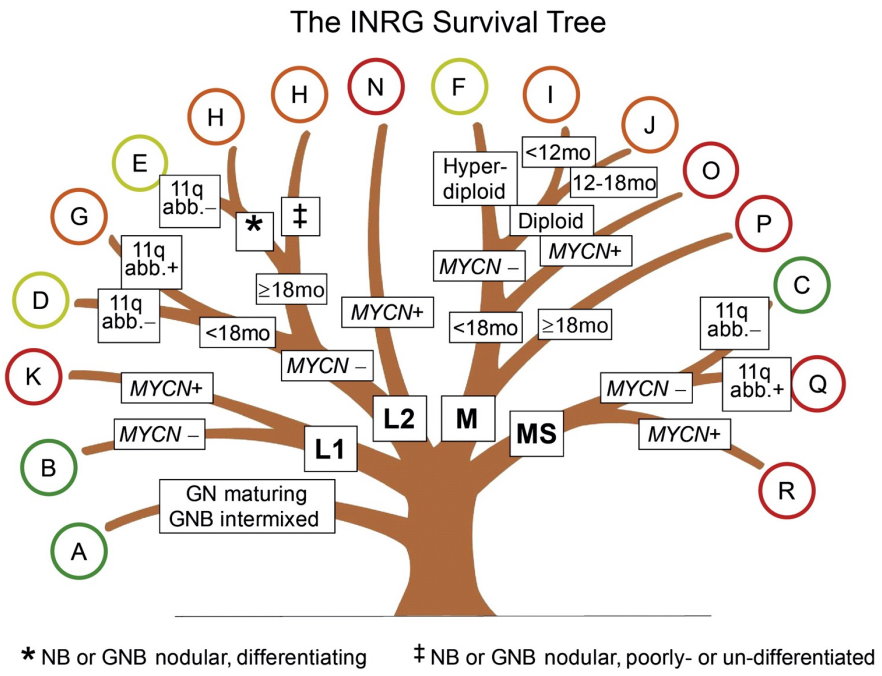

Figure courtesy of Tom Monclair, MD, PhD. Used with permission.

Figure 5. International Neuroblastoma Risk Group (INRG) Consensus Pre-treatment Classification schema. It is represented the different factors that determine the prognosis of the patient shown as branches of the tree and letters (A-R, from better prognosis to worse prognosis). L1 and L2, localized tumour; M, distant metastatic disease (except stage MS); MS, metastatic disease confined to skin, liver and/or bone marrow. GN: ganglioneuroma and GNB: ganglioneuroblastoma. MO: months. 11q: Chromosome 11. Taken from (La Quaglia, 2014). Red colours represent poor prognosis, while green colours stand for good survival-rate prognosis.

prognosis), low-risk, intermediate-risk and high-risk (bad prognosis). Likewise, non-high-risk groups (very low, low and intermediate) usually evolve positively, while high-risk patients, despite more aggressive treatments, have a lower long-term survival rate. In addition to these categories, a recent work also established an additional one: ultra-high risk (Tadeo et al., 2016), which defines a subtype of patients with extremely complicated prognosis within high-risk (Pinto et al., 2015).

A common procedure to assess NB aggressiveness is to extract a biopsy of the tumoral area to the affected patient. To be able to analyse biopsies through a microscope or a computer, they need to be prepared and stained. First, tissue preparation usually involves formalin fixation and embedding in paraffin. Paraffin blocks are then cut using a microtome (a high precision cutting instrument) and mounted on glass slides. However, most of the molecules of interest in the tissue are not readily visible or well-differentiated on the mounted sections. Therefore, the second step is the staining protocol, whose purpose is to dye the samples in order to highlight the molecules within them. Finally, a pathologist analyses the histopathological tissue sample (Veta et al.,

2014). This last step still represents the only definitive method for confirmation of presence or absence of disease, and the measurement of disease progression. Until recently, a pathologist could only analyse and visualize the tissue samples under a microscope (Titford, 2006). Currently, this workflow includes the digitalization of histopathology slides allowing the pathologist to analyse them in more detail using a computer (Gurcan et al., 2009; Veta et al., 2014). Thus, the number of stored histopathological images is increasing over time. In parallel, it is very time-consuming to thoroughly evaluate each of the images by a human expert. Altogether, it is necessary to develop new methods to objectively analyse histopathological images and, at the same time, ease pathologists' workload. Importantly, these methods will also aid the investigation on the mechanisms driving pathologies such as NB.

1.2.1 Relation between Extracellular matrix and NB

To be considered high-risk, a NB patient should either have an amplification of the MYCN oncogene in the tumour biopsy or be older than 18 months with metastatic disease (**Figure 5**). Despite the correlation between MYCN amplification and advanced staged Neuroblastomas, there are so many missing actors, aside from genetic factors. Tumours grow surrounded by non-cancer cells and their stroma, interacting with them in complex ways (**Figure 6**), forming the tumour microenvironment (TME) (Li et al., 2007). The stroma is composed of a complex mixture including the vasculature, immune cells, epithelial cells, fibroblasts, signalling molecules and the extracellular matrix (ECM). These elements give structural and connective support to cancer and non-cancer cells. In the same way, the ECM is made up of a variety of molecules supplying important structural backing for the parenchymal cells of tissues and organs (Kim et al., 2011). The ECM is mainly formed by two major classes of molecules: Glycosaminoglycans (GAGs), which has been related to signalling cascades and tissue development and homeostasis, as well as pathological processes (Afratis et al., 2012); and fibrous proteins consisting of collagen (col F), reticulin (ret F), elastin, fibronectin, and laminin. The interplay of the ECM elements is highly important to achieve tissue homeostasis. For example, collagen molecules and other fibers are dependent on the level of fluidity of the ECM that is maintained by GAGs. As hydration increases or decreases, collagen becomes more flexible or stiffer, respectively (Øien and Wiig, 2016). When stiffer, collagen fibers are more prone to be affected by changes in their mechanical microenvironment that accompanies tumour progression aiding cell migration (Ng and Brugge, 2009; Tadeo et al., 2016). Regarding NB, a stiffer ECM (cross-linked and disorganized fibers network) with an insufficient amount of collagen fibers and glycosaminoglycans, as well as irregularly shaped blood vessels, was associated to poor prognosis in NB patients (Tadeo et al.,

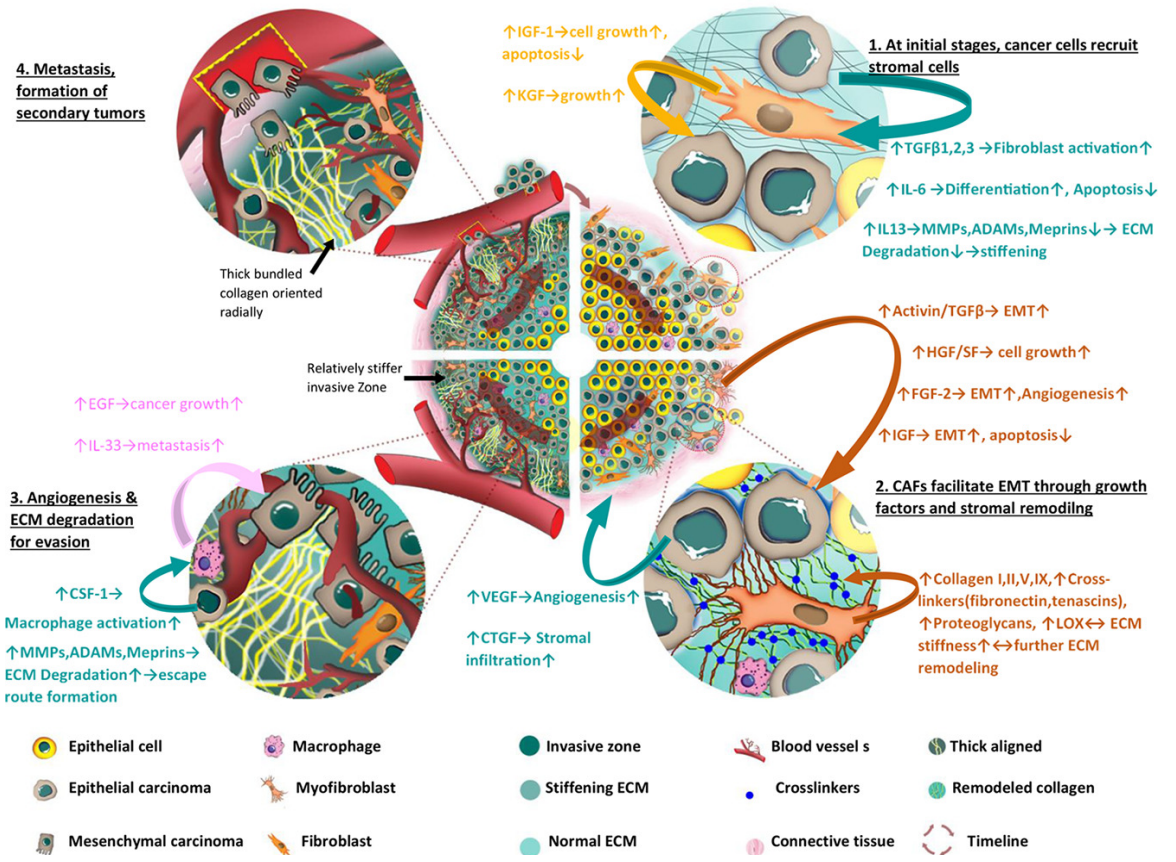

Figure 6. **Summary of the evolution of a tumour and how it modifies its microenviroment**. In particular, it shows all the steps from initial tumoral stages to metastasis, with all the biophysical and molecular remodeling ocurring on it. Taken from (Emon et al., 2018).

2016, 2017). We now know that the TME, particularly the extracellular matrix (ECM), is very relevant in cancer aggressiveness (**Figure 6**) (Gilkes et al., 2014; Pickup et al., 2014).

In parallel, the TME can be seen as a balanced compression-tension system, where cells have a central role (**Figure 7**). These notions are based on the tensional integrity (tensegrity) hypothesis that integrates how the cells are affected by its environment. Like clinicians often diagnose tumours based on differences in tissue rigidity, sensed by palpation, tensegrity links the mechanical and biochemical properties of a tissue (Ingber and Jamieson, 1985; Ingber et al., 1981). Tensegrity relies on the compendium of compression-resistant forces to balance the system. For instance, in a cellular environment, the ECM would be acting, mostly, as a compressor and, other elements, like the actin filaments within the cell, as tension structures (**Figure 7**). One of the direct consequences of this model is that our cells are physically and mechanically affecting its environment. They can, therefore, control its environment promoting ECM remodelling and, in the same way, the ECM is able to transform the way cells behave and

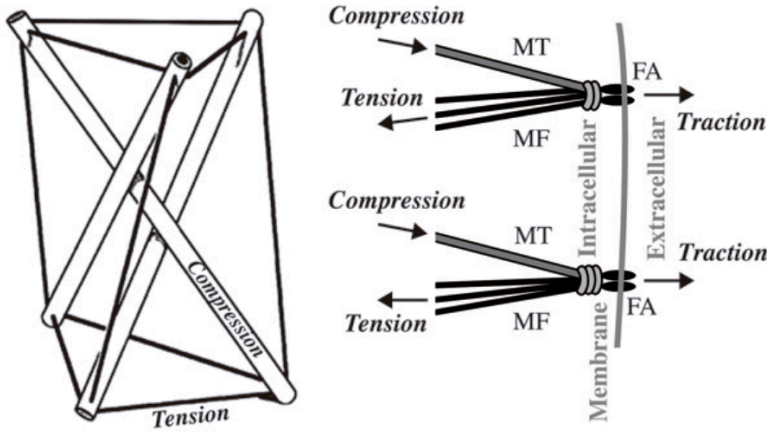

Figure 7. **Tensegrity model of a cell and its environment.** On this biophysical cellular model (Stamenović and Ingber, 2009) two main elements are at work: on the intracellular level, compression (microtubules, MT) and tension (actin microfilaments, MF); while only traction is pushing outwards on the extracellular side (traction forces at the focal adhesion, FA). As stated by the authors, the system would mechanically stabilize the cell and its shape.

function. For instance, a stiffer ECM has been shown to encourage cell migration by changing the cell traction forces (Discher et al., 2005; Lo et al., 2000).

Another essential element of the TME is the blood supply, which also works as a tensional structure of the tensegrity model (Tadeo et al., 2014). Blood vessels transport blood cells, oxygen, and nutrients to the tissues of the body. They are not only essential for cells to survive, but take an important role in tumour progression (Mandriota et al., 2001). Due to the unbalanced division cell rate of tumoral cells, tumours require more resources. To mitigate this necessity, hypoxic tumour cells recruit blood vessels by secreting vascular endothelial factors initiating tumour angiogenesis (**Figure 6**). Thus, it is commonly said that "TMEs have their own blood vessels" which are defined by its abnormal and irregular size (Nagy et al., 2009; De Palma et al., 2017).

The lymphatic vasculature, acting as a tensional structure in the tensegrity model, is critical for immune function, tissue fluid homeostasis and gastrointestinal lipid absorption. Regarding cancer disease, lymphatic vessels were assumed to be passive elements in tumour progression. Nonetheless, recent studies depict the true role of lymphatic vessels to have an integral role in the metastatic spread of disease. The presence of cancer cells around lymph nodes is associated with poor prognosis outcome in human cancer (Tadeo et al., 2018; Tuttle, 2004). In addition, the lymphatic vasculature is thought to help tumour cells to disseminate allowing the metastasis of the tumour (Mäkinen et al., 2007; Pathak et al., 2006). Thus, TME elements,

including ECM, affect NB cells in many ways: from the ability to promote angiogenesis and vascularization to enabling invasion and metastasis, among others (Borriello et al., 2016).

Overall, the tensegral hypothesis may promote different cell behaviours depending on the state of the ECM rigidity. In particular, it has been reported that the TME could be involved in a positive feedback loop between matrix stiffness, cell contractility and cell division, which could end up promoting tumour malignancy (Huang and Ingber, 2005; Ingber, 2002). In NB, there have been studies about how a stiffer ECM may correlate with poor prognostic factors (Tadeo et al., 2014, 2017). It is, therefore, timely that we further investigate this matter with the most recent tools and mathematical approaches.

In order to unveil the machinery of tumour progression, we ought to identify how, each element of the TME, affects the prognosis of NB patients. Regarding ECM elements, many of them have been already recognised as key players in cancer malignancy (Borriello et al., 2016). For instance, an ECM full of a crosslinked fibronectin and an excessive deposition of structural components such as collagen I and reticulin drive the stroma remodelling (**Figure 6**), which, utterly, would lead to metastasis (Emon et al., 2018). In NB, in particular, studies have also related crosslinking and ECM stiffness to unfavourable cancer conditions (Tadeo et al., 2017).

1.2.2 Vitronectin

In this thesis, we are going to study the possible role of one particular molecule of the ECM in NB: Vitronectin. Vitronectin (VN) is a glycoprotein linking the ECM and cells through several ligands like integrins, plasminogen activator inhibitor-1 (PAI-1) and urokinase plasminogen activator receptor (uPAR) (**Figure 8**) (Madsen and Sidenius, 2008; Preissner, 1989). It is synthesized in the liver by hepatocytes and, in a minor quantity, in the brain, lung and kidney, among others (Dimova et al., 2005; Pijuan-Thompson and Gladson, 1997; Seiffert, 1997). VN is known to preserve the vascular homeostasis (thrombosis and fibrinolysis) and controls the innate immune system by facilitating cell adhesion and migration, while aiding tissue repair and regeneration. Due to its volatile structure, the biological functions of VN are dependent on its interactions (Preissner, 1991; Preissner and Reuning, 2011). Despite the polyfunctionality of VN, the vitronectin-knockout (VN-KO) mouse was found to be viable, although the animal acquires delayed coagulation and poor wound healing (Leavesley et al., 2013).

Among the different glycoproteins of the ECM, VN has a particular interest in tumour biology. Its relationship with uPAR is associated to cell migration and signal transduction via integrins, where VN can promote cell adhesion and matrix degradation by binding to integrins, PAI-1 and

uPAR (Madsen et al., 2007). In fact, VN is considered to promote angiogenesis and vascular permeability, aiding tumour migration (Kenny et al., 2008; Li et al., 2012; Orr et al., 1992).

In NB, the role of VN remains incompletely defined, although a previous study of Dr Rosa Noguera laboratory has suggested a connection to tumour progression (Burgos-Panadero et al., 2019). Understanding the roles of VN in NB open new ways to understand better the mechanical changes in the ECM, or, in long-term, as a way to apply mechanotherapy to NB patients (Huang et al., 2013; Tadeo et al., 2014).

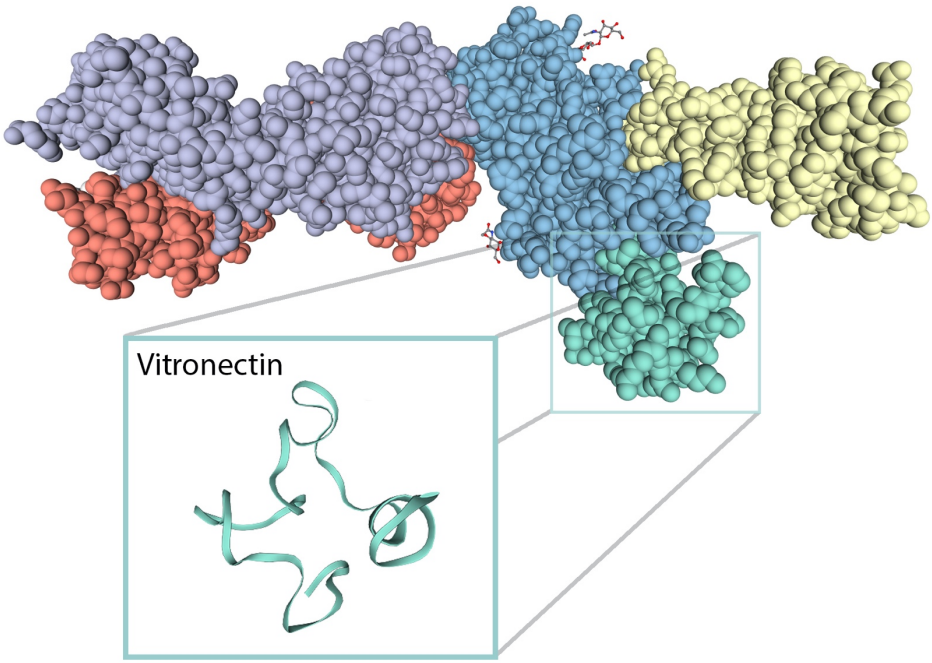

Figure 8. **Vitronectin molecular structure.** Structure of urokinase receptor, urokinase and vitronectin complex in humans (VTNC_HUMAN), and, highlighted with an aquamarine box, is the molecular structure of Vitronectin (Ribbon diagram). Each colour represents the different chain sequences: aquamarine stands for Vitronectin, Urokinase-type plasminogen activator is shown on yellow, anti-uPAR antibody is represented by two colours (light chain on malva, and heavy chain with red) and blue represents urokinase plasminogen activator surface receptor. Note that other conformations of VN are possible. Obtained from https://swissmodel.expasy.org/repository/uniprot/P04004.

2 Objectives

1) Develop new methods to improve computerized image analysis of histopathological images using network properties.

2) Quantify whether the organization of different components of the TME affects neuroblastoma malignancy using human samples.

3) Unravel Vitronectin's role in neuroblastoma tumoral environments.

4) Decipher how the topology of the tumour microenvironment affects cancer biology and its mechanisms to do so.

3 Material and methods

3.1 MATERIAL

Ninety-one primary NB tumours were acquired and processed by Rebeca Burgos-Panadero and other members of the laboratory of Dr Rosa Noguera using the protocol described at (Burgos-Panadero et al., 2019; Tadeo et al., 2016). We received mark-up images obtained from biopsy samples stained with different markers: type I collagen (Gomori staining), reticulin fibers (Masson's trichrome staining), GAGs (alcian blue), blood vessels (anti-CD31, Dako, clone JC70A, 1/50), lymphatic vessels (anti-D-2-40, clone D2-40, Monoclonal Mouse Anti-Human Podoplanin, Dako, 1/40) and VN (anti-VN, 1/100, clone EP873Y, isotype IgG, code ab45139, Abcam). In addition, we received haematoxylin stained nuclei and its unprocessed image (**Figure 9**). All this information was processed by members of the Dr Rosa Noguera's laboratory. Thus, we received segmented images, where white pixels represented areas where the molecule was found, and black pixels to the background of the image.

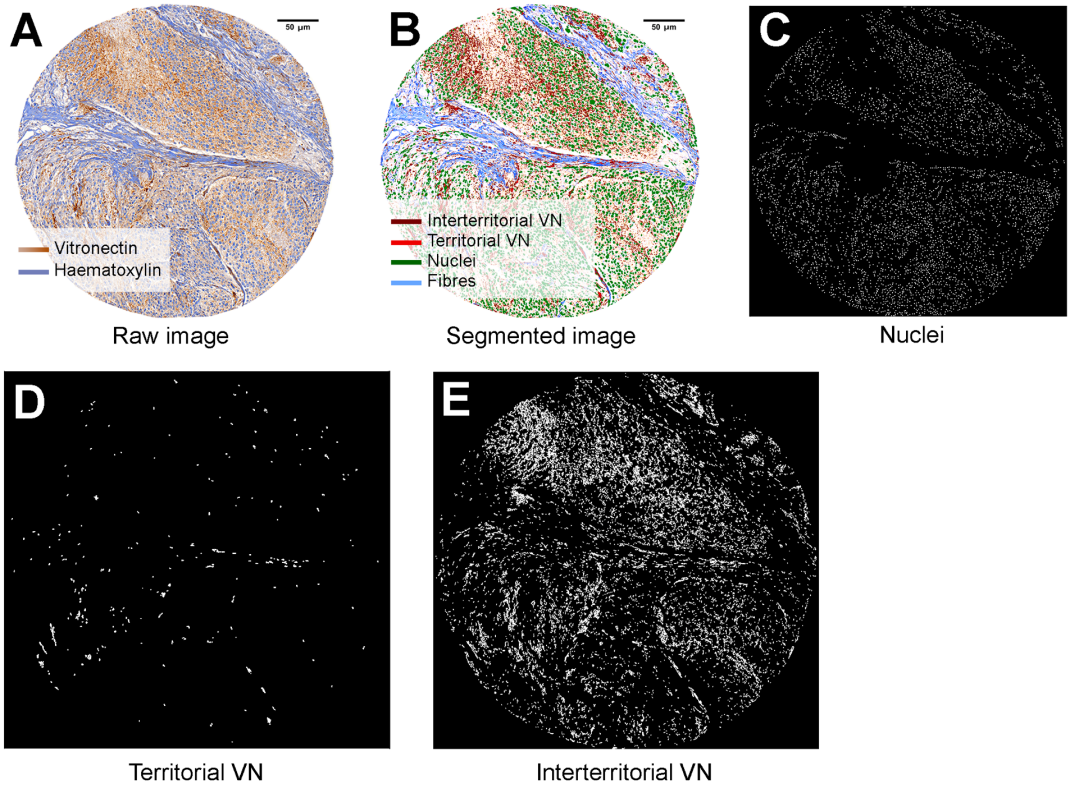

Figure 9. **Histopathological images from NB patients stained to display Vitronectin**. **A**) Represents the unprocessed biopsy image stained to visualize Vitronectin (brown colours). **B**) The processed histopathological image to differentiate between nuclei (green), interterritorial (brown) VN and Territorial VN (red). In **(A)** and **(B)**, other elements of this image are assumed to be another fibers (blue). (**B**) has been processed to obtain each of the mentioned mark-up images, where white pixels stand for the marker and the black ones are the background (**C-E**). All the images have the same resolution and size.

In addition, we got the clinico-biological information associated to each patient (and image): age≥18 months, Stage (Localized or Metastatic), Histopathologically Differentiated, 11q deletion, Ploidy (hyperdiploid, or diploid and tetraploid), MYCN amplified (MNA), segmental chromosome aberrations (SCA). Each feature was dichotomized into two categories: 0 or 1, depending on its information. Using this information, the patients were classified into high risk (medium and high risk) or low risk (low and very low risk) according to INRG system (Cohn et al., 2009). Samples were also categorized regarding its genetic instability as in (Burgos-Panadero et al., 2019): a) 'higher genetic instability' were patients with genetic profiles with >3 typical SCAs plus MNA or 11q deleted or >3 gene amplifications or with hyperrearranged chromosomal segments; b) 'lower genetic instability' were cases with genetic profiles with numerical chromosomal aberration or ≤3 typical SCAs, excluding 11q SCA.

3.2 FEATURES

To extract the features related to the processed biopsy images (**Figure 9**), we used the segmented images as input. First, we overlapped a hexagonal grid, whose hexagons had a radius of 50 pixels (approximately 8.05μm), discretizing the mark-up images into regular hexagonal areas (**Figure 10**). Each hexagonal area was a 'region', and the regions where the molecule was found became 'nodes' (see **Definitions**). The features that considered the hexagonal regions (with the suffix 'per node' or 'per region'), analysed each region of the mark-up image, separately, according to that particular feature (e.g., the Euler number of a single hexagon). Then, using both mark-up images and the information of the regions, we calculated the following characteristics related (**Figure 13**):

1. Stained area of the marker (**Figure 10**). In particular:
    i. Percentage of the stained area. We quantified the stained areas of a marker in each delimited hexagonal area and, then, we computed the mean and std of the whole grid.
    ii. In addition, we computed the mean and std of the percentage of the stained area using only the nodes (regions with stained area > 0).
2. Branches (**Figure 10**). We measured the number of crosslinks in every hexagonal region. In particular:
    i. Mean number of branches found in the image per hexagonal region (i.e. the hexagons without branches counted as 0).

ii. We also calculated the mean using only the regions with marker inside (branches per node).

3. Euler number (**Figure 10**). MATLAB's Euler number is defined as the number of objects minus the number of holes within an image. We obtained the following features:
    i. Euler number per stained area: considering the whole image, we divided the Euler number calculated for the whole image by the number of objects within the image.
    ii. We computed the Euler number considering each region, separately. We, then, obtained the Euler number per region and the Euler number per nodes, which used only the areas with information inside.
4. Holes. In particular, we acquired different measurements from the holes within the objects of the images:
    i. We measured the mean and standard deviation of the hole areas.
    ii. Number of holes per object found.
5. Only in VN, we obtained the difference between interterritorial and territorial VN. In particular, we computed the mean of the difference between interterritorial and territorial VN per region. For each hexagonal area, we operated territorial minus interterritorial.
6. Morphometry of cell nuclei (**Figure 9**). These features were acquired by the laboratory of Dr Rosa Noguera (Tadeo et al., 2016, see Table 2). They considered the whole mark-up image obtaining:
    i. The percentage of haematoxylin stained nuclei area.
    ii. Nuclei/mm^2.
    iii. Ratio of stained nuclei to total pixels.
    iv. The total nuclei number from the total mark-up images.
7. Morphometric characteristics related to the marker, also obtained by members of the Dr Rosa Noguera laboratory (Tadeo et al., 2016, see Table 2):
    i. Percentage of the stained area and number/mm^2.
    ii. Ratio of positive stained pixels to total pixels.
    iii. Positive or negative H-score (based on a specific discriminatory threshold, ranging from 0 to 300).
    iv. Number of secretory cells (percentage of positive cells).

Thereafter, we divided them into topological and non-topological (**Table 1**). In the case they capture the organization, the characteristics were labelled as topological; while features that

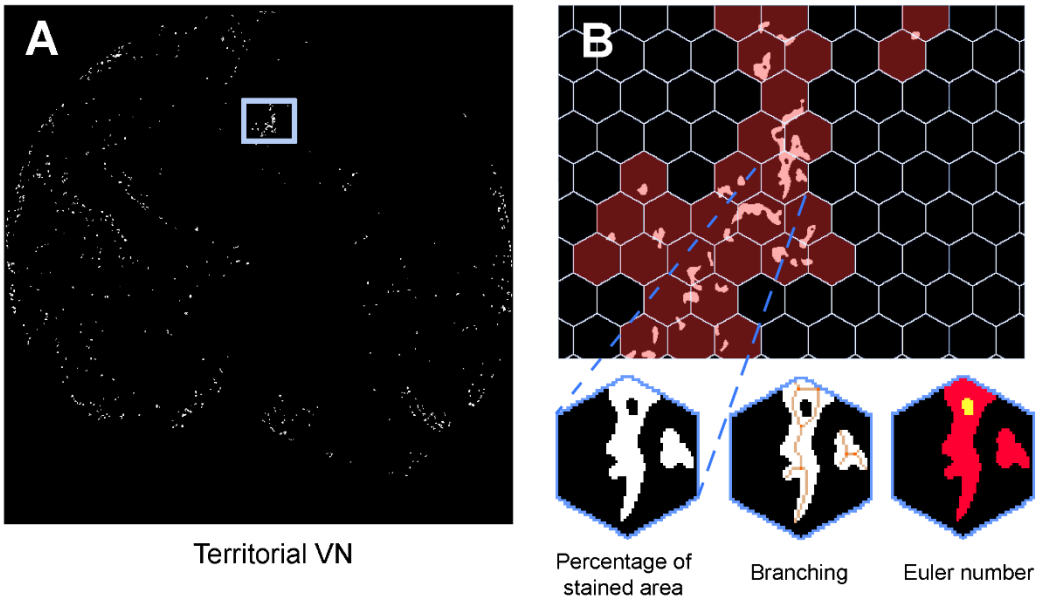

Figure 10. Examples of different features and how they are calculated. **A)** Mask image with white pixels as territorial VN and black pixels as the background. **B**) Highlighted region of **(A)** showing how the overlapping grid is discretizing that space. With a red shadow, it is presented the hexagons that have any territorial VN inside them. These hexagons will be the nodes of the network. Each node is processed independently to obtain the features of percentage of stained area, branching or Euler number, among others.

mainly consider other aspects of the image, like the morphometry of the objects or its quantity, were considered non-topological. Morphometric features obtained by Dr Rosa Noguera's laboratory were also labelled as non-topological (**Table 2**). All the features are displayed in **Table 2** (VN) and **Additional Table 1** (collagen, reticulin, GAGs, blood vessels and lymphatic vessels).

3.2.1 Pure topological features

We also added an additional set of topological features to the dataset: the topological features. Bearing in mind that nodes represent hexagonal regions with information on an image (objects of the image), we linked these nodes using the Euclidean distance between each hexagon's centre. Aiming to model how the regions are distributed throughout the sample (**Figure 10** and **Figure 13**), we computed a network for each image using three different procedures:

a) *Iteration* (**Figure 11**). Reviews all the neighbours of each node, connecting the closest ones until the network is fully connected. The algorithm follows these steps: 1) Create a distance matrix, whose rows and columns are nodes in the same order (the first row and the first column correspond to the same node). 2) Perform an iteration that involves going through all the rows of the distance matrix and 2.1) getting the closest distance of each row; 2.2) adding the column's node and row's node as edges. 2.3) Remove the edges

from the distance matrix as possible new edges for future iterations. 3) Once an iteration is over, we check if the network is connected, meaning that any hexagonal region can reach any region using the edges as a vehicle. Step 2 begins again if there are still isolated nodes.

b) *Sorting* algorithm (**Figure 11**), in contrast to *Iteration*, it does not perform an iteration *per se*. Sorting entails ordering all the distances and connecting them one at a time until the output network cannot find an unconnected region. For optimization purposes, the algorithm follows this flow: 1) Sort from closest to furthest all the distances between any pair of regions. 2) Take the smallest distance of each node from the list and obtain an array of closest distances. The largest of this array will define a threshold to add more edges on this iteration. 3) Add all the edges with a distance smaller or equal to the previously defined threshold. 4) Remove the added edges and distances from the list. 5) Finish an iteration of the algorithm. 6) Perform steps 1 to 5 until the output graph is connected. Note that this checking is performed only when an iteration is finished.

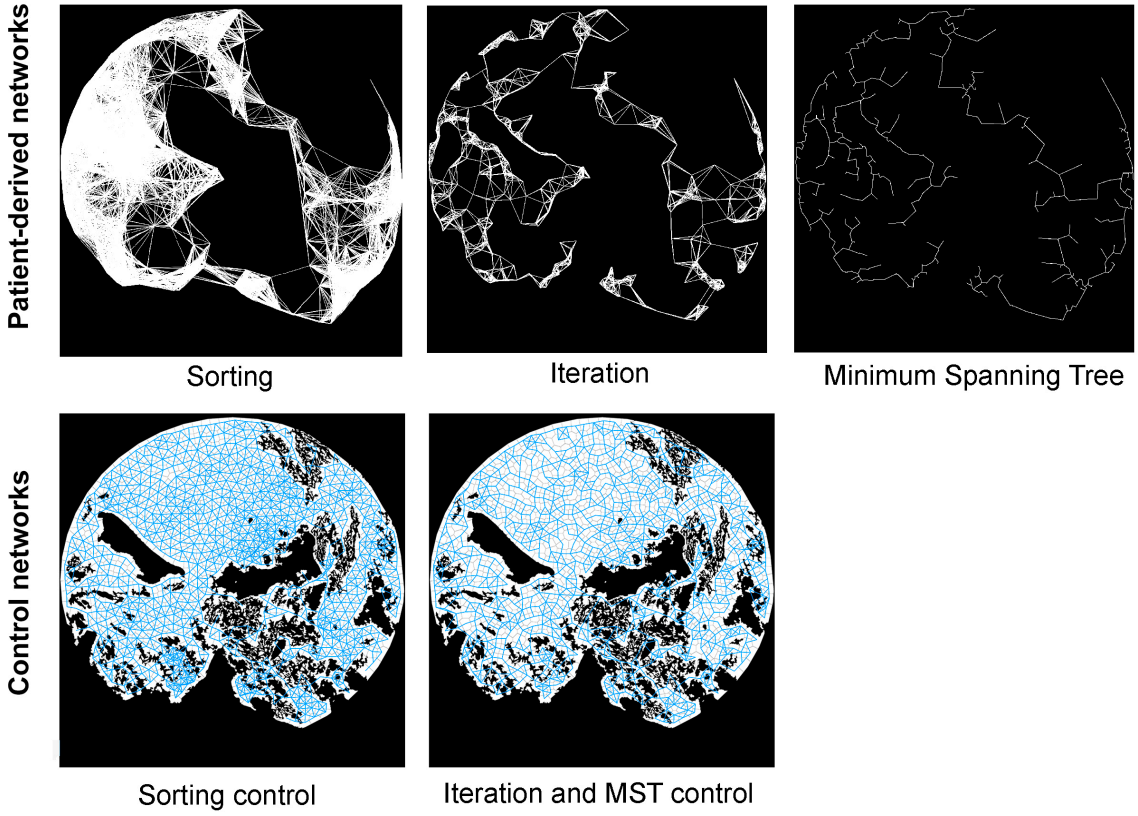

Figure 11. Networks derived from the image from Figure 10 using the three different algorithms (first row): Sorting (left), Iteration (mid) and MST (right). In addition, on the second row, it is displayed the control network of each network from the first row. On blue it is highlighted the edges of the graph, while in white it presented the control homogeneous image.

c) Minimum Spanning Tree (**Figure 11**). Although the two previous algorithms were designed on this study, the last algorithm, Minimum Spanning Tree (MST), has already been used in other studies (Yao et al., 2015). It was already implemented on a MATLAB function named *graphminspantree* (MATLAB R2014b, MathWorks). The resulting network of the MST is a network with the minimum possible number of edges, whose weights are also minimum. Moreover, this network is a tree, whose particularity is that any two vertices are connected by exactly one path without forming cycles.

Besides the networks derived from each original biopsy, we computed ten different random controls for each mark-up images of all the markers (**Figure 10**). We aimed to compare the ten homogeneous controls with the original image biopsy. The pipeline of 'control creation' follows these steps: 1) Collect the number of nodes from the original image; 2) Randomly place the same number of seeds inside an image to create a Voronoi diagram. The seeds should avoid the regions of the original image where it was not possible to find the marker (i.e. tears of the tissue). This 'valid region' was defined by the raw image, where you can perfectly delineate the shape of the biopsy. 3) Perform Lloyd's algorithm ten times, homogenizing the Voronoi cells (see **Definitions**) to become more hexagonal (Du et al., 1999; Lloyd, 1982). As a result, we obtained a control in which the regions are homogeneously spread throughout the mark-up image of the biopsy. Once the nodes were distributed, we computed the algorithm of Sorting and Iteration for the controls, whose output is a graph (**Figure 11**). In the next steps, these graphs will be compared with the tumour specific network. Iteration control network will also be used as the MST control network. In other words, we employed the Iteration control network as a control for the Iteration mark-up network and the MST mark-up network. This was due to the huge computational cost of creating an MST network from a regular pattern.

Finally, we obtained the tensegrity indices: three algorithms per marker. The tensegrity index represents the difference between the biopsies and their homogeneous controls using graphlets (**Figure 2** and **Figure 10**), which are small connected subgraphs, as a measure (Pržulj et al., 2004). Specifically, we used the ORCA (Orbit Counting Algorithm) computer program for graphlet identification and calculation (Hočevar and Demšar, 2014), to extract the different conformations of nodes assembling the graphlets, called orbits, of each network. Furthermore, we calculated the distance between two distribution of graphlets using the Graphlet Degree Distribution (GDD) distance (difference) using the 73 orbits (Przulj, 2007). We used this approach to get the GDD distance between a network of a VN location biopsy and each control. The mean of these 10

distances represented the value of the tensegrity index for that image and algorithm. The whole pipeline is explained in **Figure 13**.

3.3 STATISTICAL ANALYSIS

As a preliminary experiment, we first performed a Shapiro test to check if the tensegrity indices came from a normal distribution. Next, we performed the non-parametric Mann-Whitney test using Wilcoxon R function (Team and R Development Core Team, 2016) to test if the distribution of the tensegrity indices were statistically different regarding any of the criteria (Risk pre-classification or Genetic instability criteria) (e.g. **Table 3**).

In parallel, we computed the pipeline to obtain individual independent factors (based on previous works (Martín-Rodríguez et al., 2013, 2014) from our dataset of features (VN or TME, **Table 2** or **Additional Table 1**, respectively) that could not overlap with known predictive variables (INRG variables like age or *MYCN* status or segmental chromosome aberrations (SCA)). First, we categorized the continuous variables regarding the quartiles. In the second step, we computed a univariate analysis with each variable and calculated chi-square (χ^2), keeping only the statistically noteworthy features (**Table 4** and **Table 7**). Third, we gathered the eight characteristics with the lowest p-values (or $P<0.005$, in the case of the TME) in the high-risk pre-treatment stratification group and tumour genetic instability criteria. We, then, performed a multivariate logistic regression (**Table 8**), with a 'best subset' approach. Initially, we applied an exploratory analysis of what probabilities the selected variables have, to obtain a good subset using *mplot* (Tarr et al., 2018). Afterwards, we performed one main subset selection using *glmulti* (Calcagno and Mazancourt, 2010), and a second selection to check for the most relevant features with *bestglm* (McLeod and Xu, 2017). These relevant features were the characteristics that appeared the most on the best subsets of variables found ('% of appearance' axis, **Figure 15**, **Figure 17**, **Figure 19** and **Figure 20**). Using the most relevant variables, we obtained a model with a minimum Akaike Information Criterion (AIC) and we reported nagelkerke R^2 as a measure of the fitness of the model. The AIC represents the quality of the model relative to other models penalizing models with a higher number of coefficients, among other things (Sakamoto et al., 1987). In parallel, R^2 represents the proportional reduction in the absolute value of the log-likelihood measurement (with a log-likelihood of 1, the model fits perfectly the data, with 0 nothing at all), while a greater number of Fisher Scoring iterations is bad because it took too many iterations for the coefficients of the model to fit the data. Finally, Anova chi-sq tells you how the addition of each feature differs significantly from the expected outcome. In parallel, we measure

the proportion of true positives that are correctly identified ('sensitivity') and the true negatives that were predicted as negative ('specificity'). Using the 'VIF' function in the *car* library in R (Healy, 2005), we removed the most redundant features, retaining all the features with a value below three (we ran the function each time after removing a feature) (Petrie, 2015). We used the library of R *logistf*, which implements Firth's logistic regression (Ploner et al., 2010). Firth's method let us obtain real odds ratio by penalized profile likelihood. The results of the multivariate analysis are variables that correlate to the prognosis of NB patients or the tumour aggressiveness. Therefore, the outputs of the pipeline are individual independent factors that cannot overlap, and which add new insights to the existing ones.

3.4 CODE AVAILABILITY

All the code that has been developed for this project is available at:

https://github.com/ComplexOrganizationOfLivingMatter/NeuroblastomeIntegration

4 Results

4.1 PIPELINE OF TOPOLOGICAL ANALYSIS

This work has been done as a collaboration between the laboratory of Dr Rosa Noguera and Dr Luis M. Escudero's laboratory which became a published article (Vicente-Munuera et al., 2020). As part of a team effort, each member was assigned a different task. In particular, our group performed the computational and statistical analysis, while Dr Rosa Noguera's laboratory acquired the histopathological samples and the mark-up images coming from these samples. To sum up, our group has developed a pipeline to obtain quantitative results from the 91 biopsies that were acquired from NB human patients by the laboratory of Dr Rosa Noguera (**Material and Methods**). Each biopsy was processed to obtain serial sections that were stained to detect each one of the following molecules: reticulin, collagen, GAGs, and VN. In parallel, images of blood vessels and lymphatic vessels were also obtained. A total of 6 images per biopsy were acquired by Dr Rosa Noguera's group. Alongside these images, they obtained the haematoxylin stained nuclei mark-up images (see **Definitions** and **Figure 9**). Since we aim to focus on VN, they also distinguished two locations of VN (Burgos-Panadero et al., 2019) and separate them into two different "sub-markers" (Territorial and Interterritorial VN). They, therefore, acquired an image stained with territorial VN, whose VN is surrounding the cell (pericellular) and within them

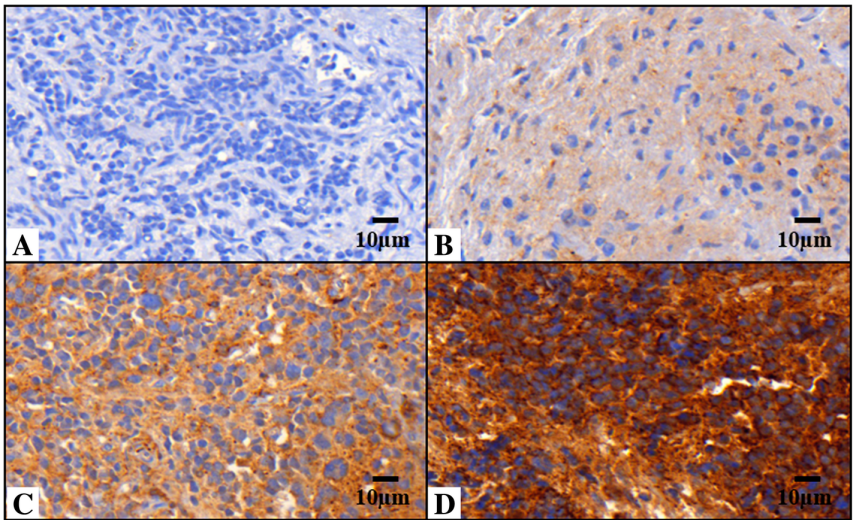

Figure 12. Vitronectin pattern in neuroblastic tumors. Images immunostained with antibody anti-vitronectin (VN) at 40X. A) Sample corresponding to negative VN. B) and C) Samples corresponding to weak and moderate VN expression and ECM distribution only (defined as interterritorial VN). D) Sample with strong VN expression with pericellular and intracellular location (defined as territorial VN). Haematoxylin is highlighted in blue, corresponding to nuclei and fibers of ECM. Taken from (Burgos-Panadero et al., 2019).

(intracellular), and another image with interterritorial VN, which is located on the intercellular space (peripherally to the territorial matrix). To obtain an objective outcome of VN staining intensity, VN's immunoreactivity was assessed by two pathologists, where it was rated as one of these four categories: no staining, and weak, moderate, and strong (**Figure 12**). Then, this classification was dichotomized as "weak to moderate" vs "strong". Based on this categorisation, an automatic image analysis was adapted to reproduce the output coming from the two pathologists (Burgos-Panadero et al., 2019). Finally, they related "strong intensity" with "territorial VN" and "weak and moderate" with "interterritorial VN". This procedure has been thoroughly explained in (Burgos-Panadero et al., 2019). Once they acquired all the images, we processed them to perform our topological analysis.

Due to the fibrous or irregular nature of all the markers, we could not capture their organization by simply detecting them as objects. To solve this issue, we divided the image into regions by overlapping a hexagonal grid of a fixed radius of 50 pixels (8.05μm) with the mark-up image (**Material and Methods** and **Discussion**). Likewise, we detected which hexagons had marker objects within them (nodes). In this way, we could use the hexagons containing the signal of a marker as a reference for the topological analysis. Using the information from the black-and-white images highlighting the markers and its associated overlapping hexagonal grid, we defined 12 features that captured the organization of the markers (Topological features, **Table 1** and **Material and Methods**). Of these 12 characteristics, three of them (Sorting, Iteration and Minimum Spanning Tree (MST) tensegral indices) were pure topological features (**Table 1**, blue light colour) representing the distribution of the marker throughout the sample biopsy. The other 9 features were related to some topological properties of the images: i) Branching (see **Definitions**): measures the crosslinks that form the molecules; ii) Euler characteristics (see **Definitions**): related to the number of the segmented objects and their inner holes; iii) deviation of the quantity of the marker within the grid and iv) properties of the holes (**Material and Methods** and **Table 1**, darker blue colour).

We also extracted 3 non-topological characteristics quantifying morphological and morphometric traits of the images (**Material and Methods** and **Table 1**, green colour). In addition, another set of non-topological features was extracted by Dr Rosa Noguera's lab (Tadeo et al., 2016, see Table 2). All the features extracted for all the markers are displayed on **Additional Table 1** (all markers except VN) and **Table 2** (VN dataset). To obtain the pure topological features, we developed a computational pipeline enabling to extract the organization of a marker in three unique ways (three tensegrity indices) using a network approach (see

Topological	Non-Topological
Name	**Name**
Sorting tensegral index	mean percentage stained area per region
Iteration tensegral index	mean percentage stained area per node
MST tensegral index	mean area of holes
std percentage stained area per region	
std percentage stained area per node	
mean quantity of branches per region	
mean quantity of branches per node	
euler number per stained area	
euler number per region	
euler number per node	
number of holes per stained area	
std area of holes	

Table 1. **List of the features extracted from the images**. In blue colour, it is presented the topological characteristics, considering that there are pure topological features (in light blue) and regular topological features (in blue). Regarding the non-topological features, they are shown on green colour.

Definitions and **Figure 13**). First, once space was discretized, we collected the centre of the hexagons that were filled with some amount of marker and identify them as nodes. Considering the Euclidean distance between the nodes, we connected them (with edges, see **Definitions**) by using three different algorithms: Sorting, Iteration and Minimum Spanning Tree (see **Material and Methods** and **Figure 11**). Second, for a given node distribution calculated from a biopsy, we computed 10 control networks with the same number of nodes. We, then, distributed them randomly, throughout the possible area it could fall into, building a new grid. Then, we homogenised the cells composing the grid using Lloyd's algorithm (**Material and Methods**). By creating a control, that considers the region within the biopsy in which the marker could be, we minimized artefacts in later experiments when comparing real and control network (**Figure 11**). Each algorithm had its own control network, except for MST. Due to the properties of the algorithm, it was impossible to create its control and we used Iteration's instead (see **Material and Methods**). Third, we calculated a similarity descriptor, based on the Graphlet Degree Distribution (GDD) distance (Pržulj et al., 2004) (**Material and Methods**), to quantify the differences in organization between the control network and the network derived from a histopathological image. For this purpose, we computed the graphlets (see **Definitions**) of each network as a measurement of their topology (**Figure 2**). The output of this pipeline was a value that we called tensegral (or tensegrity) index (**Figure 13**). Thus, we obtained three tensegral indices per marker (one per algorithm).

Topological

ID	Name
1	Interterritorial - Sorting tensegral index
2	Interterritorial - Iteration tensegral index
3	Interterritorial - MST tensegral index
4	Interterritorial - std Percentage of VN stained area per region
5	Interterritorial - std Percentage of VN stained area per node
6	Interterritorial - euler number per VN stained area
7	Interterritorial - euler number per region
8	Interterritorial - euler number per node
9	Interterritorial - number of holes per VN stained area
10	Interterritorial - std area of holes
11	Interterritorial - mean quantity of branches per region
12	Interterritorial - mean quantity of branches per node
13	Territorial - Sorting tensegral index
14	Territorial - Iteration tensegral index
15	Territorial - MST tensegral index
16	Territorial - std Percentage of VN stained area per region
17	Territorial - std Percentage of VN stained area per node
18	Territorial - euler number per VN stained area
19	Territorial - euler number per region
20	Territorial - euler number per node
21	Territorial - number of holes per VN stained area
22	Territorial - std area of holes
23	Territorial - mean quantity of branches per region
24	Territorial - mean quantity of branches per node
25	Std difference Territorial and Interterritorial

Non-Topological

ID	Name
26	Interterritorial - mean percentage of VN stained area per region
27	Interterritorial - mean percentage of VN stained area per node
28	Interterritorial - mean area of holes
29	Territorial - mean percentage of VN stained area per region
30	Territorial - mean percentage of VN stained area per node
31	Territorial - mean area of holes
32	Mean difference Territorial and Interterritorial
33	Percentage of haematoxylin stained nuclei area
34	Haematoxylin stained nuclei/mm^2
35	Interterritorial - Percentage of stained area
36	Interterritorial - VN stained area/mm^2
37	Territorial - Percentage of stained area
38	Territorial - VN stained area/mm^2
39	Total nuclei
40	Percentage of haematoxylin stained nuclei
41	Percentage of VN positive cells
42	Ratio of haematoxylin stained nuclei pixels to total pixels
43	Interterritorial - ratio of weak positive pixels to total pixels
44	Interterritorial - ratio of moderate positive pixels to total pixels
45	Territorial - ratio of strong positive pixels to total pixels
46	Ratio of all positive pixels
47	H-score

Table 2. Index of features name and identifiers used in the study, divided into topological (in blue) and non-topological (green). Topological features are the ones who capture organization, while the non-topological characteristics are morphometric measurements.

Combining all the features extracted from the 91 NB patient-derived biopsies using the above pipeline, we aim to find characteristics that could come up as new independent features of our two defined criteria: tumour genetic instability (higher vs lower, 82 cases) and pre-treatment risk stratification group (high-risk vs non-high-risk, 91 cases). In addition, depending on which type of characteristic was found relevant, we could get useful biological insights on how that marker is affecting tumour aggressiveness and progression.

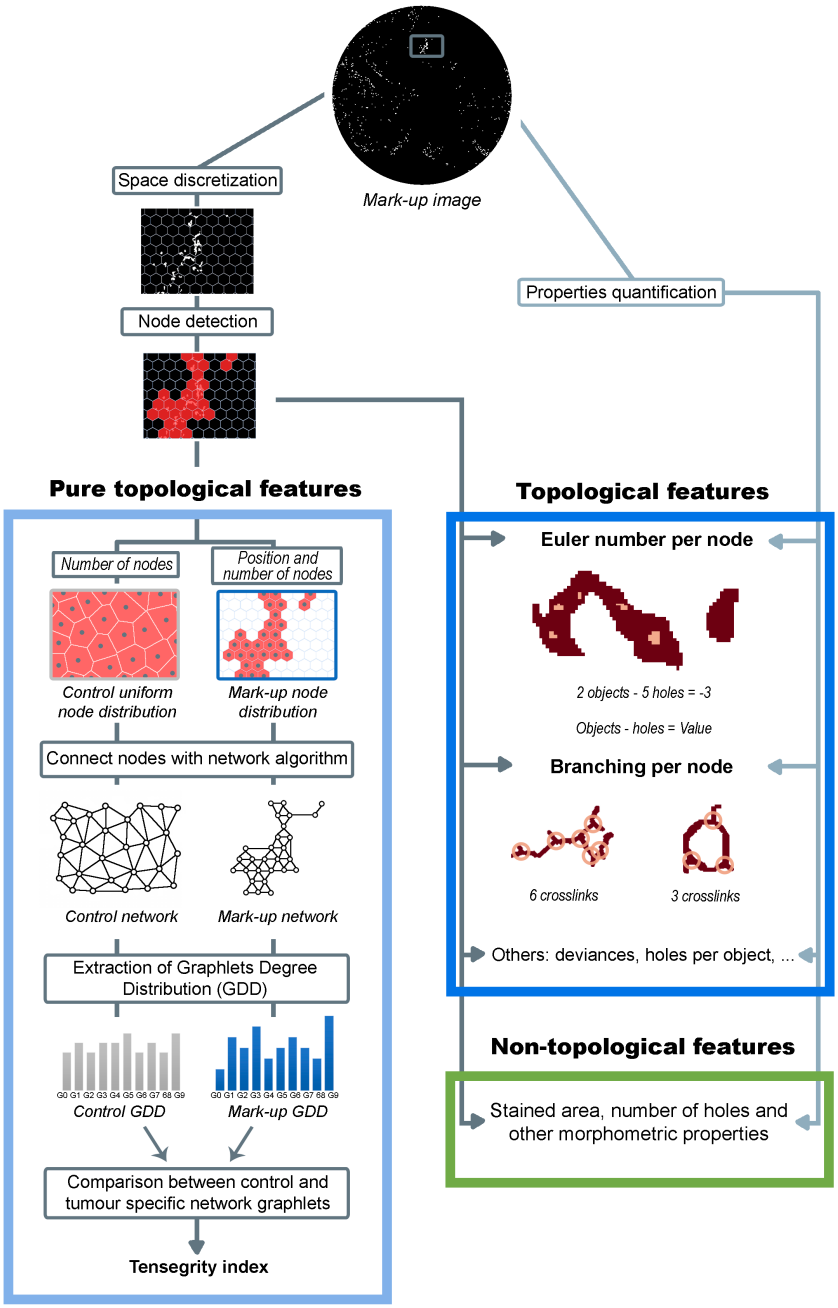

Figure 13. Pipeline overview of how the features are extracted. The process starts with the initial markup image, in which a region of interest was selected (in dark grey) to explain the space discretization. Below, the nodes (in red) are identified when a hexagon has VN inside. This information is used to obtain the pure topological features (light blue). The number of nodes is used to create the control with a uniform node distribution, while the position and number of nodes are for the node distribution. Each distribution of nodes is connected using a network algorithm and the GDD is computed for both control and markup networks. To obtain the tensegrity index, the distance between the control GDD and the markup GDD is calculated. For topological characteristics (blue), two sources of information are used: the hexagonal grid and detected nodes (arrows in darker gray), and properties quantification performed directly on the markup image (lighter gray arrows). Two topological features are highlighted: Euler number per node, where the Euler number is calculated by subtracting the two objects (in brown) against the five holes within them (in light brown); and Branches per node in which the crosslinks (circles in light brown) from territorial VN shapes (in brown) were detected. Non-topological features are in green colour.

4.2 VN ANALYSIS

VN is a glycoprotein that has been associated with cell migration and angiogenesis (Kenny et al., 2008; Madsen et al., 2007). Still, it is not well known whether this glycoprotein is implicated on NB tumorigenesis. To assess if VN could be a relevant marker for NB development, we created a VN dataset with 47 features, combining topological and non-topological features. Using the pipeline explained before (**Figure 13**), we computed 32 features from interterritorial and territorial VN (ID: 1-32), and, also, we added some features to compare them (ID: 25 and 32, **Table 2**). The remaining 15 features (ID: 33-47) were computed for a previous study (Burgos-Panadero et al., 2019) and used here.

4.2.1 Pre-analysis of pure topological features of VN

From the dataset of 47 characteristics, we created a subset of 6 to inquire whether our pure topological approach could obtain any noteworthy results related to VN distribution. Therefore, we took our 6 tensegral indices (ID: 1, 2, 3, 4, 5 and 6, **Table 2**) and checked if any of them was distributed differently between cases labelled with different prognosis or genetic instabilities (risk and genetic instability criterion).

Since our six tensegral indices did not follow a normal distribution, we computed a Mann-Whitney-Wilcoxon test to check if the two populations (within each category) were independent (**Material and Methods**). On risk criteria, we compared high-risk vs non-high-risk groups, where no feature was statistically significant between different prognoses. In contrast, considering the genetic instability criterion, we found two remarkable features that divided high and low genetic instability with statistical relevance: Iteration algorithm of interterritorial VN (lower instability: 0.22±0.08; higher instability: 0.18±0.10, $p<0.01$) and, also, Iteration tensegral index from territorial VN (lower instability: 0.26±0.08, higher instability: 0.21±0.09, $p<0.01$) (**Table 3**). More importantly, we noticed an interesting trend that related VN's topology (see **Definitions**) with the outcome of the patients: more homogenous patterns arose on high-risk group patients and higher tumour genetic instability (**Figure 14**). This is reflected in the lower values of all the tensegrity indices for both criteria, indicating a lower difference between VN's distribution and the homogeneous control (**Table 3**).

	Risk pre-treatment stratification group					
	High-Risk		*Non-High-Risk*			
	Mean	STD	Mean	STD	Shappiro	Mann-Whitney
Interterritorial VN - Sorting tensegral index	0,25	0,10	0,25	0,08	0,51	0,617
Interterritorial VN - Iteration tensegral index	0,18	0,10	0,22	0,09	1,96E-04	0,057
Interterritorial VN - MST tensegral index	0,42	0,05	0,43	0,05	3,96E-14	0,232
Territorial VN - Sorting tensegral index	0,28	0,11	0,31	0,11	3,56E-05	0,102
Territorial VN - Iteration tensegral index	0,22	0,08	0,25	0,09	7,04E-05	0,136
Territorial VN - MST tensegral index	0,40	0,12	0,42	0,10	< 2.2e-16	0,680

	Tumour genetic instability criteria					
	High		*Low*			
	Mean	STD	Mean	STD	Shappiro	Mann-Whitney
Interterritorial VN - Sorting tensegral index	0,24	0,10	0,25	0,07	0,60	0,448
Interterritorial VN - Iteration tensegral index	0,18	0,10	0,22	0,08	4,75E-04	0,006
Interterritorial VN - MST tensegral index	0,42	0,06	0,43	0,04	2,12E-13	0,114
Territorial VN - Sorting tensegral index	0,29	0,10	0,31	0,10	4,57E-05	0,103
Territorial VN - Iteration tensegral index	0,21	0,09	0,26	0,08	1,02E-04	0,006
Territorial VN - MST tensegral index	0,41	0,11	0,42	0,09	< 2.2e-16	0,498

Table 3. **Distribution of the tensegral indices regarding the patient's classifications (risk and genetic instability).** In particular, it is shown the mean and standard deviation (STD) of each category (higher and lower prognosis) with both criteria (risk and genetic instability). In yellow, there are the significant p-values (P<0.01).

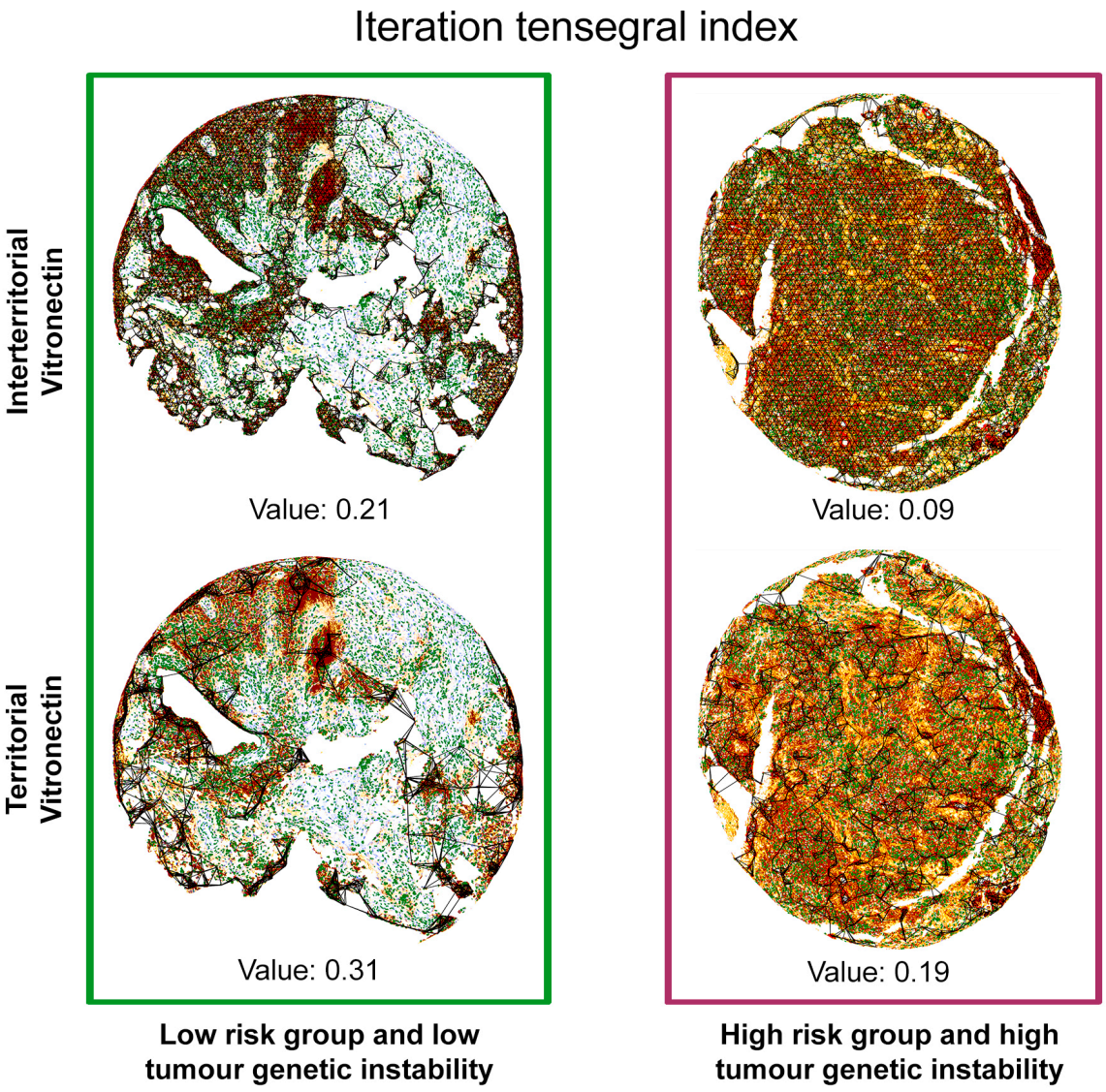

Figure 14. **Iteration tensegral index representation of different NB biopsies with different prognosis.** For the same human case, both VN locations are illustrated: interterritorial (top), where brown areas, corresponding to interterritorial VN, are connected with black links using Iteration algorithm; and territorial (bottom), in which red regions, representing territorial VN, are linked by edges creating a network. Note that images from patients related to the non-high-risk group and lower tumour genetic instability are represented in green. Burgundy rectangle shows examples of cases belong to high-risk group and higher tumour genetic instability. The 'value' is related to the Iteration tensegral index.

4.2.2 Risk group

To improve our understanding of the role of VN in NB tumoral processes, in a second step, we analysed all the VN's features (47 characteristics, **Table 2**). Considering this dataset, we looked for variables that could be related to the current patients' classification named risk pre-treatment stratification group (risk). Based on INRG category, 'risk' classifies patients into very low, low, medium and high risk depending on several tumour-tissue characteristics trying to predict its tumoral aggressiveness, and, thus, its outcome. In addition, we wondered if VN was involved in NB cancer disease in a significant way and, if so, how was affecting it from a biological perspective.

For this purpose, we made use of a statistical approach performed in previous studies (Martín-Rodríguez et al., 2013, 2014). Aiming to find a feature that could correlate with the prognosis of the patient, we used a statistical analysis based on logistic regression (**Material and Methods**). First, we searched for features that could work independently. In particular, we used a univariate logistic regression analysis. It consisted of checking if the given characteristic was able to divide high-risk patients from low risk, but only by itself. Statistically, it tested if the individual contribution of each feature is significant. The relevant features of this analysis ($P<0.05$), were labelled as independent features and the others were discarded. We encountered with 21 out of 47 (21/47) variables that were independent for risk classification group, where 11 of them (11/21) belonged to territorial VN (**Table 4**).

The next step was to carry out the multivariate analysis consisting of obtaining the most relevant characteristics for a given criterion. Therefore, a selected variable on this step would mean that it could become a potential marker to assess one of the criteria. First, we acquired the 8^{th} most independent features with the lowest p-value (but took 9 because the last variables had the same p-value). Since we aim to find independent variables, we put together the selected features from the previous univariate analysis and the INRG clinico-biological parameters, which are recognized predictor variables (**Material and Methods** and **Table 5**). We, then, checked how this model behaved (**Table 5**).

Risk pre-treatment stratification group

ID	Characteristics	Chi-sq
20	**Territorial - euler number per node***	**0,001**
24	**Territorial - mean quantity of branches per node***	**0,003**
39	Total nuclei	0,003
37	**Territorial - Percentage of stained area***	0,006
16	**Territorial - std Percentage of VN stained area per region**	**0,010**
22	**Territorial - std area of holes**	**0,010**
34	Haematoxylin stained nuclei/mm^2	0,010
36	Interterritorial - VN stained area/mm^2	0,010
45	**Territorial - ratio of strong positive pixels to total pixels**	**0,010**
43	Interterritorial - ratio of weak positive pixels to total pixels	0,014
17	Interterritorial - std Percentage of VN stained area per node	0,016
29	**Territorial - mean percentage of VN stained area per region**	**0,019**
23	**Territorial - mean quantity of branches per region**	**0,019**
19	**Territorial - euler number per region**	**0,019**
47	H-score	0,019
27	**Territorial - std Percentage of VN stained area per node**	**0,019**
38	**Territorial - VN stained area/mm^2**	**0,019**
33	Percentage of haematoxylin stained nuclei area	0,028
42	Ratio of haematoxylin stained nuclei pixels to total pixels	0,028
30	**Territorial - mean percentage of VN stained area per node**	**0,033**
46	Ratio of all positive pixels	0,033

Table 4. Results from the univariate analysis performed for Risk pre-treatment stratificiation group. Only statistically significant characteristics (chi-sq<0.05) are shown. The features are ranked by their p-values obtained on the chi-square test, in ascending order. The selected features to be used in the next steps appeared with darker colours. The characteristics of territorial vitronectin (VN) are marked in bold (12/21 in the risk group). Highly statistically significant common features in risk group were marked with an asterisk. Using the same scheme of colours from **Table 2**. MST, minimum spanning tree; std, standard deviation.

To compare two statistical models, we reported the Akaike Information Criterion (AIC, **Table 5**). Therefore, we always prefer a model with minimum AIC value and we reported nagelkerke R^2 as a measure of the fitness of the model (**Table 5** and **Material and Methods**). Considering these statistical hints, we found this model too much explicative (R^2 nagelkerke: 0.65), with a number of Fisher Scoring iterations higher than it should (14) and, more importantly, many of its independent variables did not explain (significantly) the dependent variable (risk) (**Table 5**). In addition, the model was yet not optimal (p-value: 0.01). Thus, it was required to remove some features. To achieve the best model, we computed the 10 best subset models in terms of AIC and see which features appeared most. Using two types of 'best subset' algorithms ('Bestglm'

and 'Glmulti', **Material and Methods)** and an additional exploratory algorithm, we selected 5 variables that were the most important to achieve a remarkable model (**Figure 15**, in orange: Age, Stage, SCA and MYCN; and in blue: Euler number of Territorial VN). Once we obtained this model, we looked for collinearities. A collinearity is observed when two (or more) features can be represented, approximately, with the same regression line. That means that the collinear characteristics would be explaining the same results and they would be only adding noise. Collinearities are solved by getting only one of the redundant features. We found collinearities between SCA and MYCN because their predictions were overlapping. Therefore, we solved this issue by removing the 'SCA' variable from the model because it obtained a greater value of collinearity. The resulting final model had 4 features, with one of our topological characteristics (**Figure 15** and **Table 6**).

	Estimate	Standard error	p-value
(Intercept)	-3.43	1.37	**0.01**
16: Territorial VN – std percentage of VN stained area per region	-1.36	1.10	0.21
24: Territorial VN – Mean quantity of branches per node	1.12	0.94	0.23
20: Territorial VN – Euler number per node	1.71	0.94	**0.07**
22: Territorial – std area of holes	0.16	0.49	0.75
34: VTN - Haematoxylin stained nuclei/mm^2	-0.25	0.40	0.54
36: Interterritorial VN – V stained area/mm^2	0.39	0.53	0.45
37: Territorial VN - Percentage of stained area	0.60	0.87	0.49
39: VN - Total nuclei	-0.18	0.74	0.80
45: Territorial VN - Ratio of Strong pixels to total pixels	-1.52	0.92	**0.10**
Age (≥18 month)	3.53	0.88	**5.00E-05**
Stage	-0.01	0.01	0.66
Histologically differentiated	0.00	0.00	0.73
SCA	-1.93	0.84	**0.02**
MYCN (MNA)	1.92	0.84	**0.02**
Ploidy	0.00	0.00	0.46
11q deletion	0.00	0.02	0.84

Akaike Information Criterion (AIC): 88.75
Number of Fisher Scoring iterations: 14

Nagelkerke R^2: 0.65
Anova chi-square: 0.01

Table 5. Logistic model result after univariate analysis in terms of Risk classification. On the left column are represented the features used on the model and the intercept, which helps to adjust the model. For each feature, it is displayed the estimate (log odd ratio), std error (standard error), equivalent to the significance of the estimated coefficient (p-value or Pr(>|z|) (**Material and Methods**). We have divided with a dotted line the following group of variables: the intercept, the features that came from the univariate analysis and the clinico-biological variables, which are known predictors of NB risk and genetic instability criteria. On the bottom, it is shown the properties used to evaluate the logistic models (AIC, fisher scoring iterations, Nagelkerke R^2 and anova chi-square). The variables with P<0.1 are displayed in bold.

	Estimate	Standard error	p-value
(Intercept)	-3.94	0.91	**1.00E-05**
20: Territorial VN – Euler number per node	0.65	0.26	**0.01**
Age (≥18 month)	2.66	0.61	**1.00E-05**
Stage	-0.01	0.01	0.61
MYCN	-0.01	0.01	0.62

AIC: 84.31
Number of Fisher Scoring iterations: 8

Nagelkerke R^2: 0.47
Anova chi-square: 0.001

Table 6. Final model regarding risk pre-treatment stratification group, composed by 4 features: Euler number of Territorial VN (a topological feature of our dataset), Age, Stage and MYCN. As in **Table 5**, it is displayed the information for each variable and the properties to evaluate the model.

Comparing this model (**Table 6**) with the previous model (**Table 5**), we got that this final model is quite better regarding our reported parameters (AIC: 84.31 vs previous AIC: 88.75, due to the lower number of independent variables; chi-sq: 0.001 vs previous chi-sq: 0.01; and a fair number of fisher iterations). Likewise, this model obtained a specificity of 0.89 (non-high-risk group) and a sensitivity of 0.74 (high-risk group) (**Material and Methods**).

In terms of the independent features composing the model regarding the risk group of the patient, we analysed the relative importance of each variable within the model. To do that, we used the log odd ratio represented in **Table 6**. A log odds ratio is obtained by computing the logarithmic of the odds ratio. To easily interpret the log odds ratio, we needed to see if it was lower or higher than 0, significantly. In particular, If the log odds ratio of a particular variable is > 0 (e.g. patient's age), being all the other variables equal, we obtained that as the given feature increases its value, the patient is more likely to have a higher instability. In the case of an odd ratio < 0, a lower value in the feature points out to a higher tumour-specific genetic instability. Note that all the independent variables have an associated p-value, which tells you if the feature changes, it will probably change the genetic instability of the patient using this particular model. In the case of an odd ratio very close to 0, the p-value will not be statistically significant (P>0.05) as in 'Stage' and 'MYCN' features, which tells us that the most predictive features of the model are the Euler number per node of territorial VN (P<0.05) and the age of the patient (P<0.001) (**Table 6**).

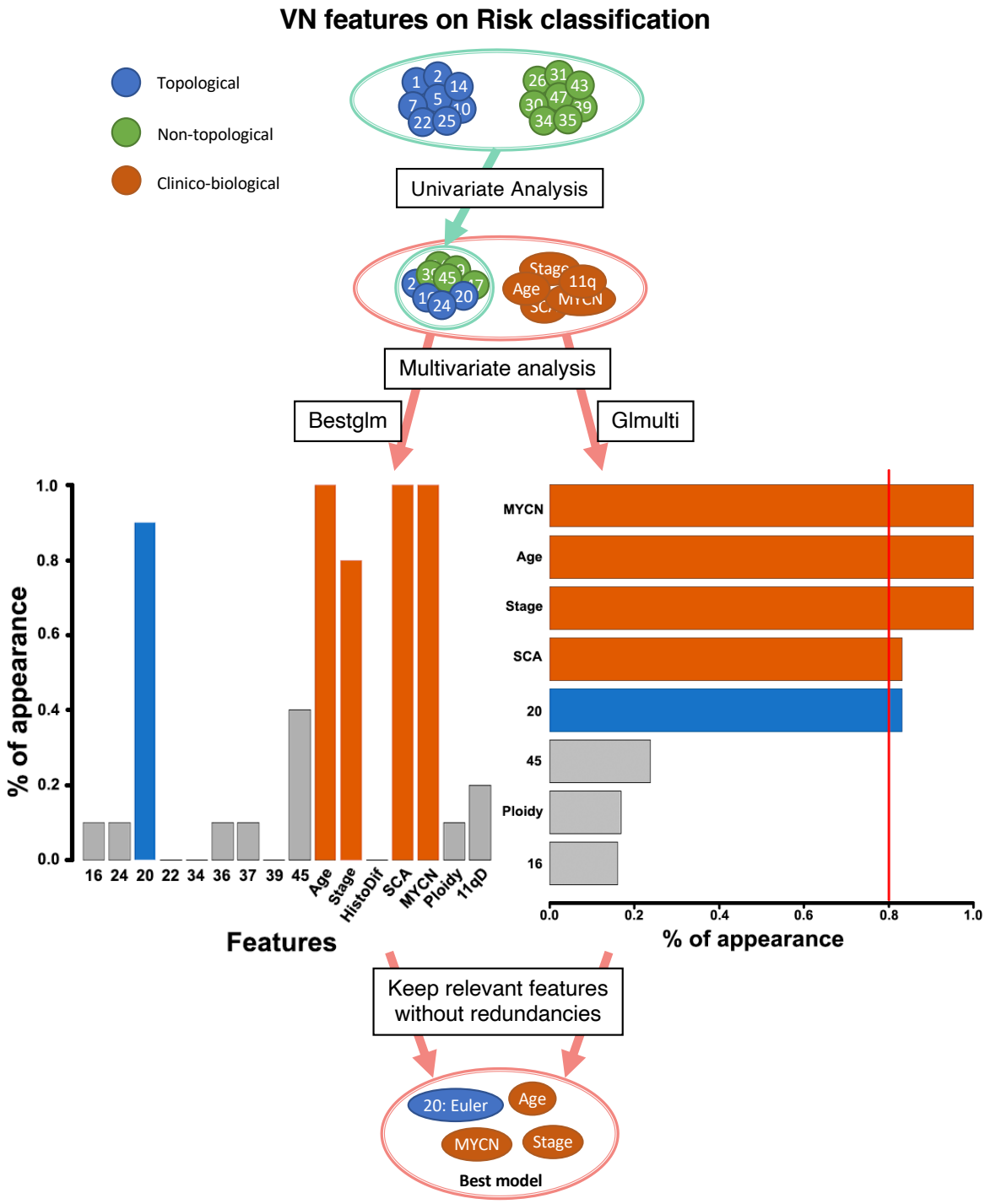

Figure 15. Statistical pipeline with VN dataset to obtain the most relevant features for Risk classification. Starting with the 47 VN features (topological in blue and non-topological in green colour), we performed a univariate analysis getting the 'independent' variables. With the 9th most statistically independent and the clinico-biological dataset (orange), we run a multivariate analysis with two different algorithms (Bestglm and Glmulti, **Material and Methods**). Finally, after keeping only the relevant features and removing redundancies, we obtained the best model with 4 features: 20 (Euler), Age, MYCN and Stage. Note that variables in grey were not selected for following analyses.

Overall, we have added a new independent variable based on the Euler number for territorial VN that is positively correlated with the risk of the cases, i.e. as the value of this feature increases, the risk does it accordingly (**Figure 16**). The Euler number stands for the number of objects in a sample minus the number of holes within those objects (**Material and Methods**). However, it may take into account more aspects of VN topology and structure. The feature Euler number per node from territorial VN not only is related to the quantity of pericellular VN but considers the compactness of the territorial VN stained area including intracellular VN. Finally, we found that a higher Euler number per node from territorial VN was associated with the high-risk pre-treatment stratification group.

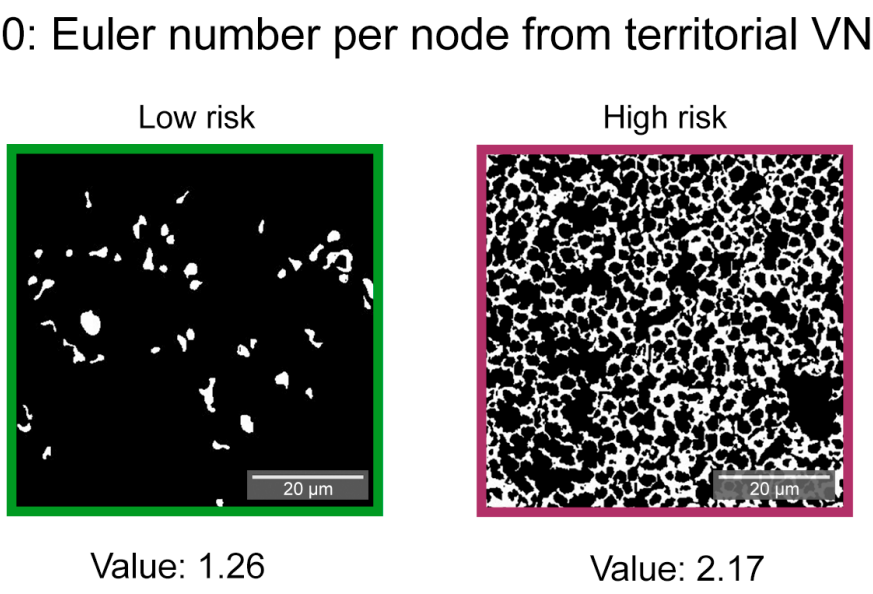

Figure 16. **Two examples of the Euler number on a low risk patient (left, green) and a high risk one (right, gaunt)**. In green, there is a region of interest of a non-high-risk patient whose biopsy obtained a value of 1.26 on the feature Euler number per node from territorial VN. On gaunt, it is highlighted a selected area with Euler number of 2.17 for an image coming from a NB patient.

4.2.3 Genetic instability criterion

Following the same pipeline as before, we first performed a univariate analysis considering the patient's genetic instability criterion. Among the resultant independent features, the three most statistically significant in tumour genetic instability were related to territorial VN features, also presenting low chi-sq values (**Table 7**). As in risk classification analysis happened, the majority of possible independent factors were characteristics of territorial VN (**Table 4** and **Table 7**, green rows).

Tumour genetic instability criteria

ID	Characteristics	Chi-sq
20	**Territorial - euler number per node***	**8,20E-07**
24	**Territorial - mean quantity of branches per node***	**4,03E-06**
37	**Territorial - Percentage of stained area***	**6,58E-05**
17	Interterritorial - std Percentage of VN stained area per region	9,05E-05
39	Total nuclei	1,27E-04
29	**Territorial - mean percentage of VN stained area per node**	**1,75E-04**
16	**Territorial - std Percentage of VN stained area per region**	**2,74E-04**
45	**Territorial - ratio of strong positive pixels to total pixels**	**2,74E-04**
27	**Territorial - mean percentage of VN stained area per region**	**3,87E-04**
23	**Territorial - mean quantity of branches per region**	**3,87E-04**
7	**Territorial - euler number per region**	**5,34E-04**
40	Percentage of haematoxylin stained nuclei area	6,11E-04
22	**Territorial - std area of holes**	**8,44E-04**
43	Interterritorial - ratio of weak positive pixels to total pixels	1,46E-03
47	H-score	0,002
38	**Territorial - VN stained area/mm^2**	**0,002**
46	Ratio of all positive pixels	0,004
31	**Territorial - mean area of holes**	**0,005**
42	Ratio of haematoxylin stained nuclei pixels to total pixels	0,006
36	Interterritorial - VN stained area/mm^2	0,009
12	Interterritorial - mean quantity of branches per node	0,014
34	Haematoxylin stained nuclei/mm^2	0,016
2	Interterritorial - Iteration tensegral index	0,017
19	**Territorial - euler number per region**	**0,024**
21	**Territorial - number of holes per VN stained area**	**0,030**
9	Interterritorial - number of holes per VN stained area	0,034
6	Interterritorial - euler number per VN stained area	0,045

Table 7. **Selected features from the univariate analysis for tumour genetic instability criteria (P<0.05)**. Rows are color-coded regarding its topological basis as in **Table 2**.

We selected the 8th most relevant features that presented the lowest p-values on the univariate analysis (**Table 7**). Next, we computed a logistic regression coupling the previous subset of characteristics and the clinico-biological parameters of the patients (**Table 8**). Still, this model was too explicative (Nagelkerke R^2: 1.00) and not optimal due to a high number of variables. Therefore, as performed for risk pre-treatment stratification group, we carried out a multivariate analysis to obtain which variables were the most relevant independent factors for the genetic instability criteria. Diverse features appeared on the selected subset of the two 'best-subset

	Estimate	Standard error	p-value
(Intercept)	-2,00E+02	1,10E+05	1,00
17: Interterritorial VN - std Percentage of VN stained area per region	-9,00E+01	2,60E+05	1
29: Territorial VN - mean percentage of VN stained area per node	3,80E+01	1,60E+05	1
16: Territorial VN - std percentage of VN stained area per region	-5,20E+01	1,00E+05	1
24: Territorial VN - mean quantity of branches per node	7,10E+01	4,80E+04	0,999
20: Territorial VN - Euler number per node	7,90E+01	1,10E+05	0,999
37: Territorial VN - Percentage of stained area	-2,00E+01	2,00E+05	1
39: Total nuclei	-4,50E+00	1,10E+05	1
45: Territorial VN - ratio of strong positive pixels to total pixels	1,10E+01	3,20E+04	1
Age (≥18 month)	8,30E+01	8,10E+04	0,999
Stage	1,10E-01	3,40E+02	1
Histologically differentiated	-7,30E-03	8,10E+01	1
SCA	1,00E+02	1,30E+05	0,999
MYCN (MNA)	9,40E+01	1,20E+05	0,999
Ploidy	-5,00E-02	1,00E+02	1
11q deletion	-4,70E-02	2,50E+02	1

AIC: 32 **Nagelkerke R^2:** 1.00
Number of Fisher Scoring iterations: 25 **Anova chi-square:** 0.0003

Table 8. Resulting logistic regression model of Risk stratification group using the intercept (top), independent features (middle) and the clinico-biological variables of the patients (bottom). We displayed, for all the variables, its estimate log odd ratio, standard error, and p-value. To analyse the model, it is displayed its corresponding AIC, Fisher scoring, R^2 and chi-sq.

selection' algorithms (**Material and Methods**). However, since the 'Glmulti' algorithm is more accurate (**Figure 17**), we only took the ones that appeared more than 80% on this method and were relevant on 'Bestglm' (appearance>=0.5) (**Figure 17**), obtaining the variables that were the most noteworthy. Therefore, we discarded features "11q deletion" and 17, because the former had a lower prevalence on 'Bestglm' algorithm (**Figure 17**), and the latter was not selected by the 'Glmulti' algorithm (**Figure 17**).

Despite we have used a robust method, other issues might occur. We, then, looked for collinearities, and we discovered that Stage was collinear to feature 24 (Territorial VN – mean quantity of branches per node). We, therefore, removed Stage because it obtained worse results than feature 24 on the multivariate analysis (**Figure 17**). Thus, our best model assessing tumour genetic instability was formed by genetic profile (SCA), *MYCN* status (MNA), ploidy (diploid and tetraploid) and mean quantity of branches per node of territorial VN (**Figure 17**). With this set of four variables, we obtained the most optimal model (**Table 9**, regular logistic regression with AIC: 41.99, number of Fisher scoring iterations: 19, chi-sq: 4.03e-06, Nagelkerke R^2: 0.84). Although we obtained the most relevant independent factors from our dataset, we acknowledged the issue of quasi-completion separation to acquire the real value of the odds ratio. Mainly, this problem is defined by extreme odds ratio with enormous standard errors.

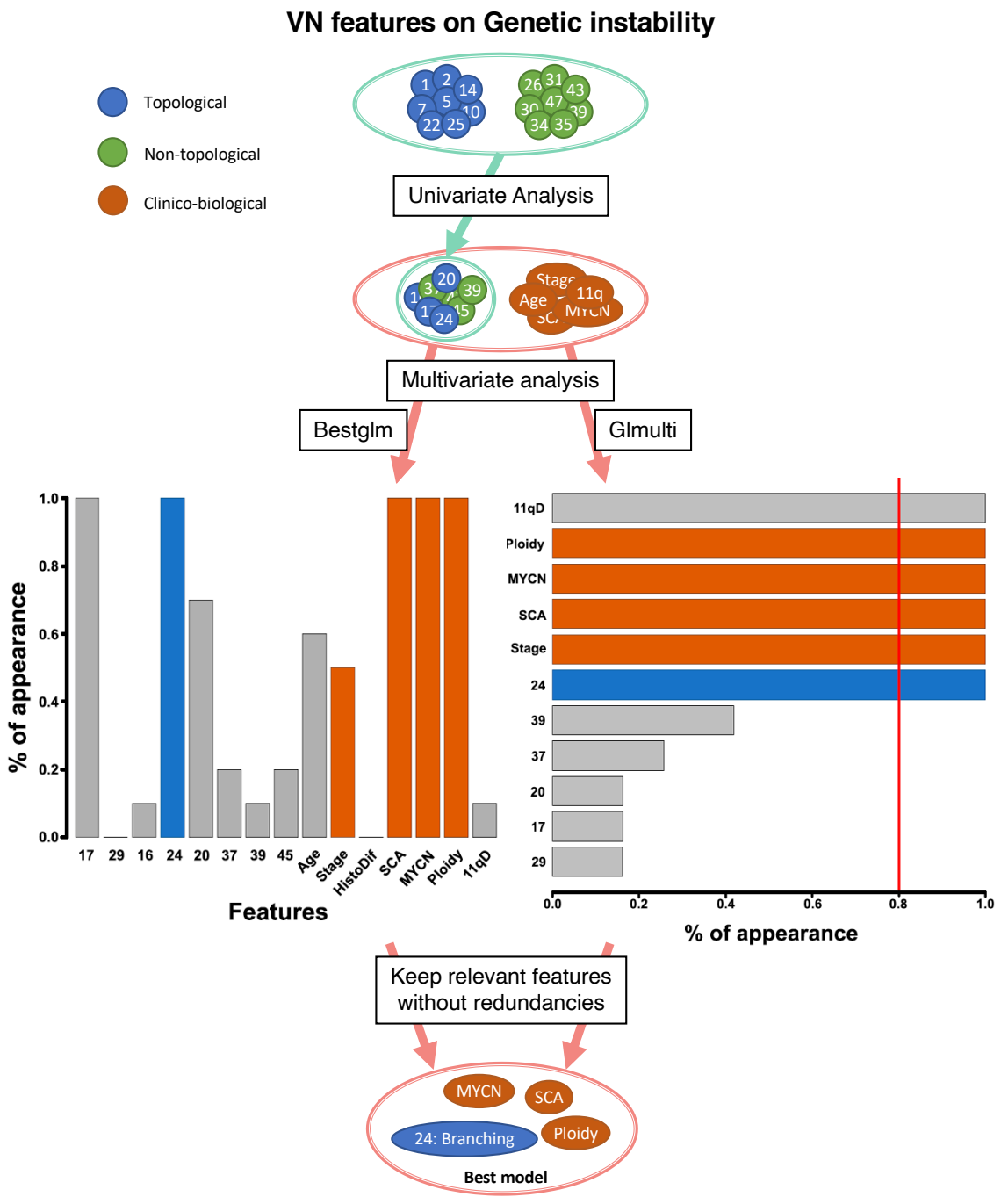

Figure 17. Statistical protocol to find the most relevant features of VN in terms of the genetic instability of NB patients. Using the VN dataset (topological: blue, non-topological: green, both: aquamarine colour), we computed a univariate analysis to obtain the 8th best features (**Material and Methods**) assessing the Genetic Instability independently (aquamarine ellipse). We, then, combined the resultant features from the univariate analysis with the clinico-biological characteristics (orange coloured). With this dataset, we performed a multivariate analysis with two methods (**Material and Methods**). By removing redundancies, we kept only the relevant features for this criterion. Grey variables were not selected as relevant for this classification.

To avoid this common issue (Allison, 2004; Heinze and Schemper, 2002), we performed a penalized model called Firth's logistic regression (**Material and Methods**), instead of using the default logistic regression (**Table 9**). Using this penalized model (χ^2<0.005), we obtained a specificity of 0.91 and a sensitivity of 0.89.

Regular logistic regression

Variable	B	S.E	Exp (B) (95% CI)	z-value	Pr(>|z|)
(Intercept)	-23.64	2914.00	5.45E-11 (0 - Inf)	-0.01	0.994
Territorial - Mean quantity of branches per node	1.50	0.58	4.46 (1.44 - 13.80)	2.60	0.009
SCA	19.89	2914.00	4.37E+08 (0 - Inf)	0.01	0.995
MYCN (MNA)	22.58	3245.00	6.43E+09(0 - Inf)	0.01	0.994
Ploidy	-2.83E-03	1.32E-03	1.00 (0.99 -1.00)	-2.15	0.032

Firth's logistic regression

Variable	B	S.E	Exp (B) (95% CI)	Chi-sq	p
(Intercept)	-6.53	2.10	1.42E-04 (4.58E-06 - 4.64E-02)	21.99	2.73E-06
Territorial - Mean quantity of branches per node	1.24	0.47	3.45 (1.42 - 10.96)	8.00	0.005
SCA	3.45	1.58	31.45 (3.04 - 4.45)	10.20	0.001
MYCN (MNA)	5.26	1.95	192.31 (9.81 - 4764.41)	20.04	7.56E-06
Ploidy	-2.21E-03	1.04E-03	1.00 (1.00 - 1.00)	5.46	0.019

Table 9. **The two final models of genetic instability criteria**. The table located on top represents the regular logistic regression with huge standard errors. In contrast, the bottom table stands for the penalized Firth's logistic regression with its corrected values. Both models share the same features, with different associated log odds ratio (B), standard error (S.E.), odds ratio and its confidence score (exp(B) 95%CI), anova chi-square value (chi-sq) and p-value (or z-value and Pr(>|z|).

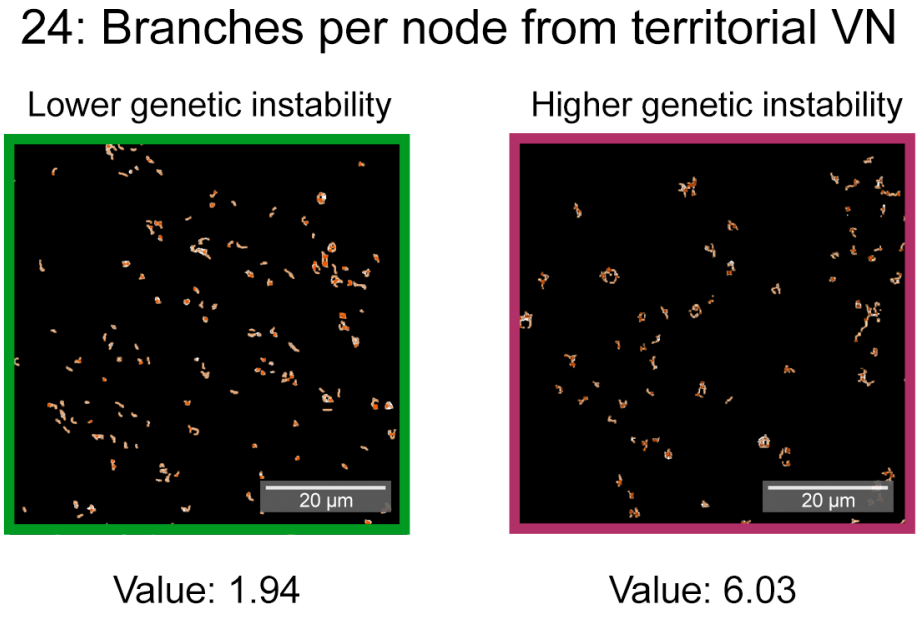

Figure 18. Branches per node of Territorial VN feature. It is illustrated, two region of interest examples with different genetic instability: on green colour (left), the example of an area taken from a lower genetic instability; on garnet colour (right), a region of interest representing a higher value of the branching correlated with a higher genetic instability.

Interestingly, one of our variables calculated for this work appeared on the final model of genetic instability: the mean quantity of branches per node of territorial VN. In particular, this feature considers the number of crosslinks after skeletonizing the image, taking into account only the hexagons of the grid with some territorial VN (**Material and Methods**). Thanks to the Firth's logistic regression, we obtained the true value of odds ratio of each variable alongside with its importance (p-value) assessing the tumour genetic instability criteria. For the selected branching feature, we obtained a positive correlation between the branching and the tumour genetic instability of the patient, meaning that the higher the number of branches, the worse the tumour genetic instability (**Figure 18**). Importantly, we found a trend on the four defined levels from the genetic instability of the tumours (very low: 1.81±0.8; low: 2.30±1.6; medium: 2.43±1.1; high: 7.05±5.4). Overall, we can say that territorial VN crosslinks is highly correlated to the tumour genetic instability.

4.3 THE TUMOUR MICROENVIRONMENT

To quantify whether the organization of the components forming the TME affects neuroblastoma malignancy, we performed the same approach used with VN in previous sections. We acquired a set of topological and non-topological features using different markers of the TME from the same biopsies and patients. We analysed and extracted pure topological, topological, non-topological characteristics from the following TME elements: collagen, reticulin, GAGs, blood vessels and lymphatic vessels. We, then, created an additional dataset formed by all these features (114 in total, **Additional Table 1**) from each marker to weight the relative importance of VN in assessing cancer aggressiveness and tumour genetic instability.

4.3.1 Risk group

The first step was to create a subset with all the features, but excluding VN, (**Additional Table 1**) to find the relevant features within the TME dataset. Following the statistical pipeline (**Material and Methods**), we only kept the characteristics that obtained $P<0.005$ for the risk classification on the univariate analysis. This subset was formed by three features related with the blood vessels (ID: 78, 95 and 96), two with the collagen (108 and 110) and one with the reticulin (ID: 72, **Table 10**). To perform the multivariate analysis, we added the clinico-biological features and the VN features that came from the past analysis (Branching and Euler number of territorial VN). The protocol and results of this analysis are shown in **Figure 19**.

	Estimate	Standard error	p-value
(Intercept)	-4.41	3.35	0.19
78: Blood vessels – std percentage stained area per node	-0.49	0.48	0.31
108: Collagen - mean quantity of branches per node	0.19	0.6	0.78
110: Collagen - euler number per region	0.59	0.54	0.27
72: Reticulin - Dendrites	0.05	0.41	0.90
95: Blood vessels - PerRatio	-0.91	0.41	**0.03**
96: Blood vessels - Perimeter	0.76	0.49	0.12
Age (≥18 month)	3.12	0.98	**1.40E-03**
Stage	-0.01	0.01	0.64
Histologically differentiated	0.00	0.00	0.68
SCA	-1.70	0.98	**0.08**
MYCN (MNA)	1.70	0.97	**0.08**
Ploidy	0.00	0.00	0.51
11q deletion	-0.01	0.02	0.71
20: Territorial VN – Euler number per node	1.02	0.82	0.21
24: Territorial VN – mean quantity of branches per node	-0.20	0.77	0.80

AIC: 82.5
Number of Fisher Scoring iterations: 14
Nagelkerke R^2: 0.69
Anova chi-square: 0.003

Table 10. Logistic regression model with all the TME variables (including VN) for risk pre-stratification group. It is formed by the variables presented on the left column and, for each feature, is shown its estimate, and associated standard error, the p-value. With a dotted line it is divided the different groups of features: the intercept, the features coming from the univariate analysis, the clinic biological characteristics and the known relevant features of VN from previous analysis. In addition, it is presented the goodness of the model in the bold properties below (AIC, fisher scoring iterations, R^2 and chi-square).

	Estimate	Standard error	p-value
(Intercept)	-2.39	1.22	**0.05**
95: Blood vessels - PerRatio	-0.52	0.30	**0.09**
Age (≥18 month)	2.31	0.64	**3.00E-04**
Stage	-0.01	0.01	0.62
MYCN (MNA)	0.00	0.01	0.69
20: Territorial VN – Euler number per node	0.58	0.27	**0.03**

AIC: 83.301
Number of Fisher Scoring iterations: 8
Nagelkerke R^2: 0.50
Anova chi-square: 3.2e-05

Table 11. Final model of VN and the rest of TME elements for risk criterion. It is represented all the information as in **Table 10**.

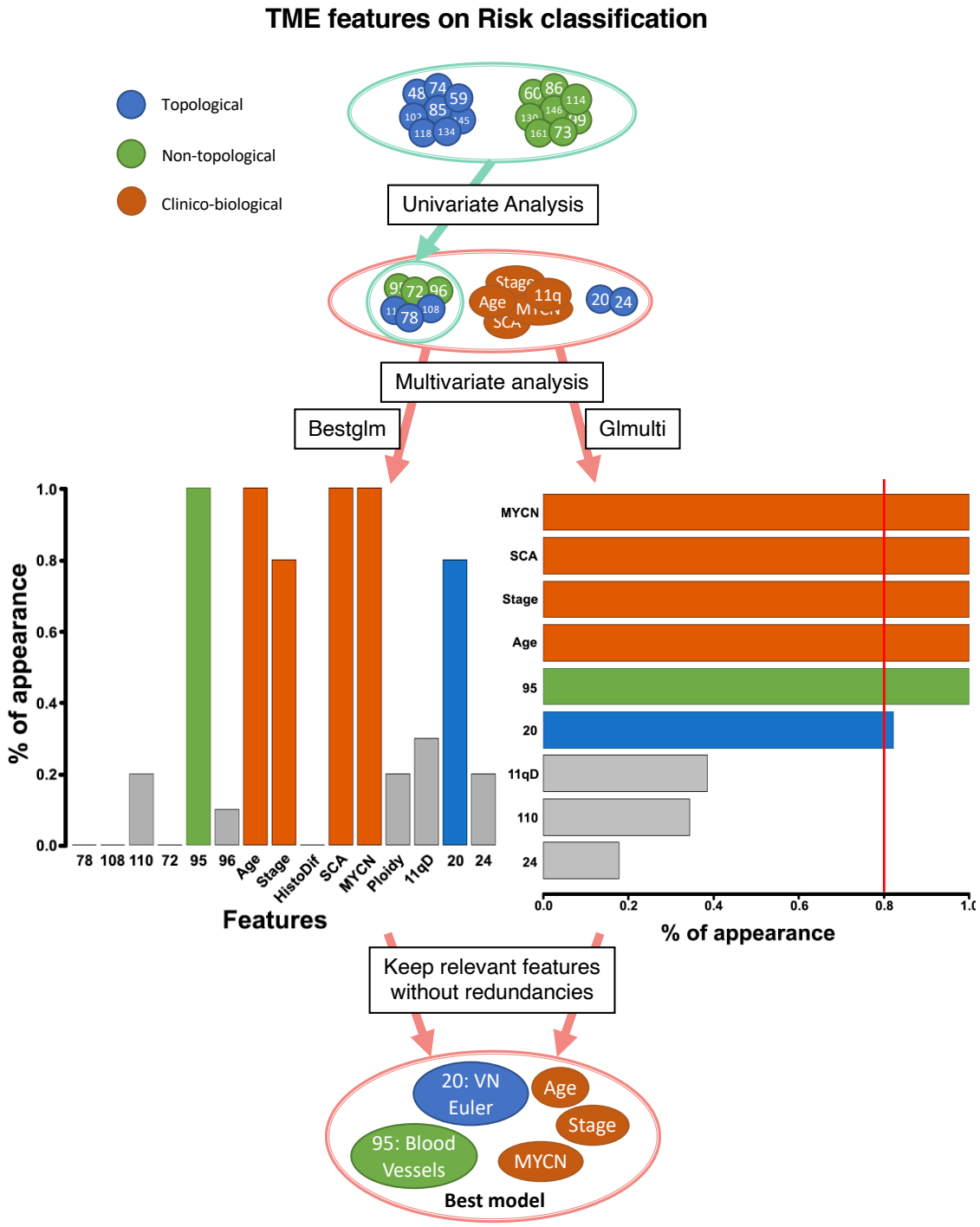

Figure 19. Statistical analysis using TME dataset regarding Risk classification. The initial dataset was formed by variables from collagen, reticulin, blood vessels, lymphatic vessels and GAGs (i.e. TME features). We obtained the most significant variables working independently (P<0.005) using a univariate analysis. Then, we coupled the previous subset, the clinico-biological variables and the VN features found to be important (**Table 3** and **Table 6**). Considering these features, we performed a multivariate analysis and after removing redundancies, we obtained the following best model: 20 from Territorial VN, 95 of blood vessels, Age, Stage and MYCN. Note that each colour represents the category of the features: green (non-topological), blue (topological) and orange (clinico-biological). Variables in grey colour were found to be not relevant for this criterion.

The best model we obtained from the logistic regression is shown in **Table 11**. Interestingly, the Euler number of territorial VN was still a relevant feature and appeared on the final model alongside the other clinical characteristics (**Table 6**). Likewise, a blood vessel feature was included in the final model. Specifically, the aspect ratio of the blood vessels became an individual independent feature for the risk classification of NB patients. The results suggest that more regular blood vessels would be associated with lower risk patients.

4.3.2 Genetic instability criterion

Following the same criteria used for Risk classification, we acquired 4 features (P<0.005) resulting from the univariate analysis performed (ID: 108, 67, 72, 95, **Table 12**). We, then, joined these variables with the known predictors and computed the multivariate analysis to reduce the number of variables of the model. Despite we have added new possible features from reticulin or collagen, the output of the analysis is the same than the obtained in VN analysis: the only feature that was found relevant is the "Mean quantity of branches per node" of Territorial VN (ID: 24, **Figure 20**).

	Estimate	Standard error	p-value
(Intercept)	-1.55E+15	5.38E+07	**<2e-16**
108: Collagen - mean quantity of branches per node	-2.38E+14	8.09E+06	**<2e-16**
67: Reticulin - Roundness	1.05E+14	1.18E+07	**<2e-16**
72: Reticulin - Dendrites	-5.77E+14	1.22E+07	**<2e-16**
95: Blood vessels - PerRatio	-1.08E+14	7.84E+06	**<2e-16**
Age (≥18 month)	1.08E+15	1.81E+07	**<2e-16**
Stage	-6.35E+11	7.15E+04	**<2e-16**
Histologically differentiated	-6.54E+11	3.43E+04	**<2e-16**
SCA	2.69E+15	1.89E+07	**<2e-16**
MYCN (MNA)	1.58E+15	2.07E+07	**<2e-16**
Ploidy	-1.13E+12	1.67E+04	**<2e-16**
11q deletion	-2.33E+12	5.16E+04	**<2e-16**
20: Territorial VN – Euler number per node	1.57E+14	1.27E+07	**<2e-16**
24: Territorial VN – mean quantity of branches per node	1.74E+14	1.29E+07	**<2e-16**

AIC: 316
Number of Fisher Scoring iterations: 19
Nagelkerke R^2: -10.11
Anova chi-square: 0.0002

Table 12. Initial logistic regression model using the independent variables from TME dataset regarding Genetic Instability. The variables included came from the univariate analysis and the known predictors (INRG variables and both territorial VN characteristics), divided by dotted lines. It is displayed all the names of the features along with its estimate coefficient, standard error and its p-value. Denoted by all the statistically significant p-values, it is not a good model (**Material and Methods**). It is also shown the AIC, number of Fisher scoring iterations, R^2 and chi-square to demonstrate the goodness of the model.

In conclusion, for instability, we find no suitable addition to the model beyond what we had (clinico-biological variables and VN mean quantity of branches). Therefore, results shown in **Table 9**, are still valid.

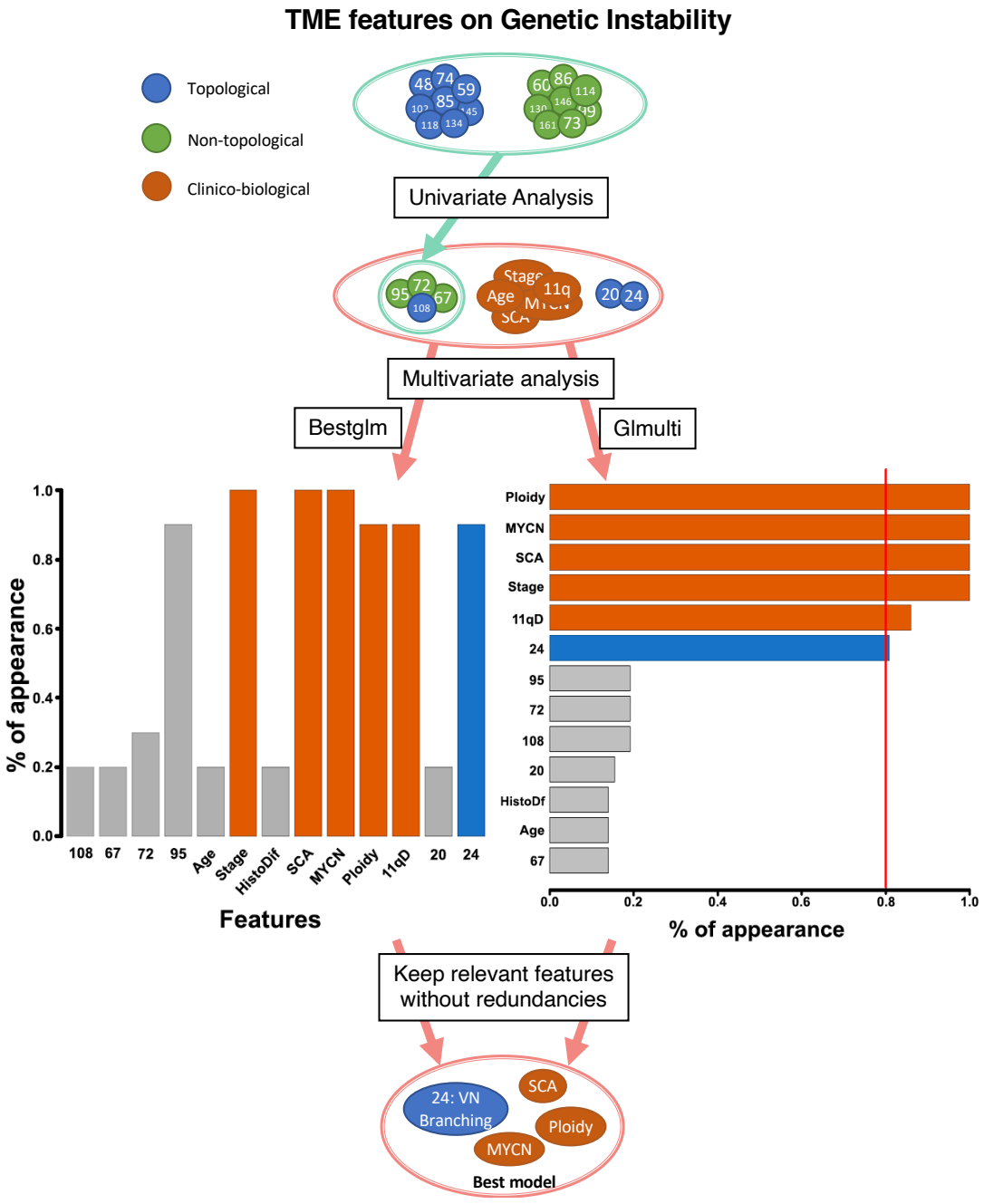

Figure 20. Statistical protocol to obtain the most relevant subset of TME dataset regarding Genetic instability criterion. We, first, performed a univariate analysis to the TME dataset, acquiring the most statistically independent characteristics (P<0.005). Then, we combined the resulting subset (aquamarine ellipse) with the clinico-biological features and the two selected VN variables of previous analysis. Using two algorithms (Bestglm and Glmulti, **Material and Methods**), we computed a multivariate analysis, in which we kept the most relevant features. By removing redundancies, we obtained the best subset of variables to assess Genetic instability of NB patients.

5 Discussion

In this work, we show that the organization of VN can be useful to assess the genetic instability and prognosis (risk) of NB patients. Using the set images of (Burgos-Panadero et al., 2019), we perform a computational approach to characterise VN and other TME elements, in terms of organization and morphology, aiming to better understand how these elements affect tumour progression. Overall, we find that, within our dataset, the VN organization is a relevant feature for both tumour genetic instability and risk classification of NB patients. In addition, our data suggest a plausible scenario in which the way VN is organized could play a greater role than previously thought, regarding tumour aggressiveness.

Networks are nowadays a common approach in biology (Vicente-Munuera et al., 2020; Viola et al., 2019). However, there are many unexploited tools that can be still used. In this study, we developed a complex pipeline (**Figure 13**) to obtain objective information about how is structured a marker (e.g. VN or Collagen) throughout a histopathological NB image, by means of the graphlets of its network. For this purpose, we computed the GDD (see **Abbreviations**) distance between the mark-up image network and its uniform control for a particular marker and algorithm, resulting in a simple number that reflects how similar are the two graphs. Therefore, the GDD will capture how the marker is spread throughout the biopsy (its difference with the homogenous control). Considering that the graphlets of the controls will be enriched in the same graphlets due to its hexagonal (and homogeneous) distribution, a marker, accumulated differently through the sample, would be enriched by distinct motifs, thus, presenting a higher tensegral index (that goes from 0 to 1).

Due to the simplicity of the resulting values acquired in the pipeline of the pure topological features (**Figure 13**), we wanted to minimize the technical artefacts that could arise from the discretization we used. For that reason, we tested a range of hexagonal sizes for the image discretization (radius from 3 pixels to 300 pixels; 0.483μm to 48.3μm, respectively). Since the size of each region is directly related to the number of nodes that would be in the calculated network, it would affect the results of the pure topological features. On one hand, if we have big regions, there would be only a few nodes in the network. Our graphlets algorithm (GDD) is meant to work with large and complex networks, where there are different motifs and in diverse quantities (Przulj, 2007). On the other hand, smaller hexagons would result in a greater quantity of nodes causing a more enriched network and graphlets. However, it would be highly dependent on the quantity of the marker, instead of its organization. After several tests, we selected the 50

pixels hexagonal radius (8.05μm, **Figure 10**) as an intermediate size, optimal to capture how the analysed markers were organized locally and globally.

Moreover, to compute the network of an image from its hexagonal discretization, we use three different algorithms (Iteration, Sorting and MST). Considering these outcomes, we have found that the Iteration algorithm was the most informative (**Table 3**, **Table 4** and **Table 7**). We interpret that networks computed with the Iteration algorithm will be more suitable to be characterized by graphlets. These results could be related to the very dense networks (enriched by the same graphlets) provided by Sorting algorithm, while MST networks, with a very low density of edges (MST networks are tree graphs, i.e. they have no cycles), would be enriched by "line graphlets". (**Figure 2** and **Figure 11**). These graphlets configurations would make the GDD less sensitive to differences with the control and, therefore, less powerful to discriminate the different criteria (Risk and Genetic instability).

Focusing on the VN analysis, although the tensegral indices were not selected as the most relevant features, they do reveal that VN behaves differently in cases with a worse prognosis. Our results indicate that more homogeneous patterns of VN appear when the tumour is more aggressive (**Table 3**). However, we reckon that these results might be reflecting not only its organization but the quantity of VN in the biopsy. In a particular scenario where VN is all over the biopsy more homogenous pattern would arise. Still, we have shown that graphlets may be useful to characterise different organizations, even when both images resemble similar to the naked eye (**Figure 21**). Future works will be needed to continue exploring the utility of graph theory algorithms to capture organizational information from biopsies.

To assess how the topology of VN was related to tumour aggressiveness and the prognosis of the patient, we extracted morphological and topological features from NB patients with different prognosis. We obtained two topological characteristics that correlated with each criterion: Euler number of territorial VN with Risk pre-treatment classification; and branching of territorial VN with Genetic instability criterion. In the first analysis, we found that the Euler number of biopsies taken from patients with worse prognosis was, on average, higher (**Figure 16**). It may be found difficult to biologically interpret this feature due to its versatility and complexity. First of all, we think that the quantity of VN is affecting the Euler number. For instance, a direct consequence of our grid discretization is that if the quantity of VN increases, there would be more objects and fewer holes. Still, how VN is structured and interconnected with all the elements of the tissue is considered. For instance, **Figure 16** might be illustrating how, in a NB patient with lower risk, the VN is surrounding the cells and within them (**Figure 16**, left). However, a different scenario is

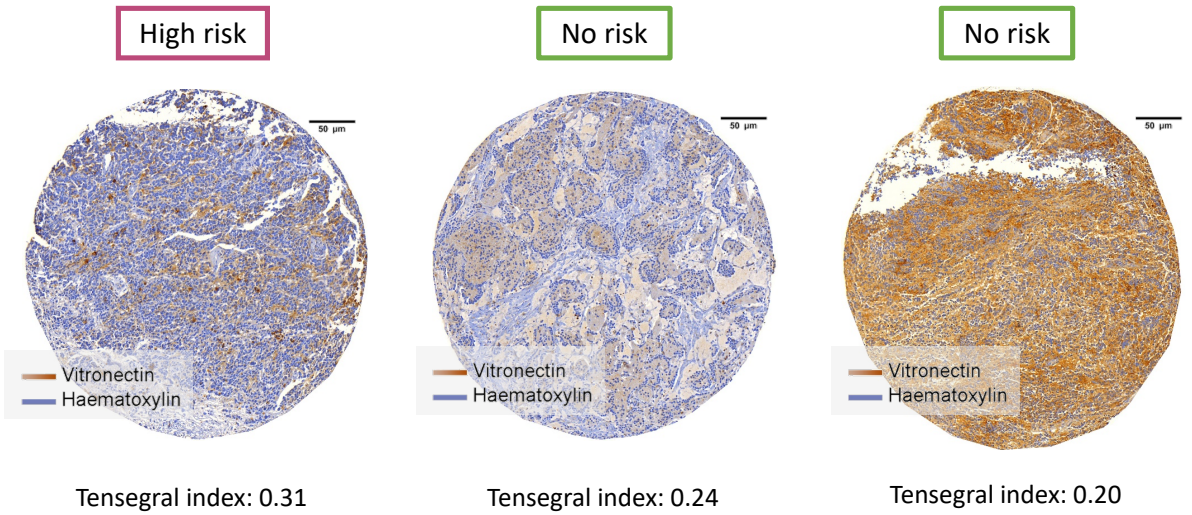

Figure 21. Examples of biopsies with different organizations and prognosis are shown. Biopsies are stained to display VN (orange-brown colour) and haematoxylin (blue colour). The risk classification associated to each biopsy is also displayed. Tensegral indices correspond to Iteration tensegral indices of Territorial VN.

represented when the tumour is more aggressive. On **Figure 16** (right), VN might be more required (thus, a higher quantity would also be presented) exhorting the cell to produce it, in a disorganized way, showing a 'holey' pattern representing a higher Euler number. A similar pattern arose on a different work, where prostate histological images were analysed (Wittke et al., 2007). They found that tubular and cribriform growth patterns were associated with lower and higher aggressive behaviours, respectively. Similarly, we encounter that in territorial VN, a higher Euler number is related to worse prognosis (**Figure 16**). Thus, considering our results, territorial VN may be forming cribriform-like structures that would surround the cancer cells and, as a consequence, remodelling the ECM.

For the tumour genetic instability criteria, the branches per node of territorial VN was the most relevant feature regarding our statistical analysis (**Figure 17**). In particular, we found that the higher number of branches, the more aggressive the tumour would be. We suggest that this feature could also be related to the shape of VN formed in the surroundings of the cells (**Figure 18**). Putting together all the results, we conjectured why the ECM and VN could be interplaying in terms of tumour aggressiveness. Altogether, considering the biological insights from the two most relevant features (Euler number and Branching of Territorial VN), we propose that VN may be creating migration tracks, aiding the tumour to progress to other organs, which, in turn, would decrease the survival possibilities of NB patients (**Figure 22**). This metastatic process on NB

would involve a disturbing change on the structure of the ECM, in which VN would be, at least, implicated.

In tissue homeostasis, the concentration of territorial VN, located closer to the cell or within it, and interterritorial VN must be regulated by a steady proteolytic activity. Imbalances in proteases were found to control tissue homeostasis by degrading the ECM, which, in turn, helps the tumour to invade and metastasizes (Kessenbrock et al., 2015; Sevenich and Joyce, 2014; Zhang et al., 2012). VN has already been associated with defects in proteolysis regulation and cell adhesion (De Lorenzi et al., 2016). In our case, membrane-associated proteases, either overexpressed or by its lack, could be the cause of an over proliferated territorial VN due to a malfunction of pericellular proteolysis (Toromanov et al., 2015). Another possibility is that endothelial cells are migrating to a fibronectin-rich environment as a way to create new blood vessels in the process of neoangiogenesis (DeClerck et al., 2004; Isogai et al., 2001). It also concords with the homogeneous patterns found in both territorial and interterritorial VN (**Figure 14**), which were spread without limits in poor prognosis NB patients. Even more, both possibilities may be in agreement and could be joining forces (Kraniak et al., 2018): a higher presence of territorial VN would change dramatically the constitution and mechanics of the ECM by the rapid addition of new synthesized VN creating migration tracks, that may lead to a more aggressive NB (**Figure 22**). However, it is unclear which is the mechanism behind these results.

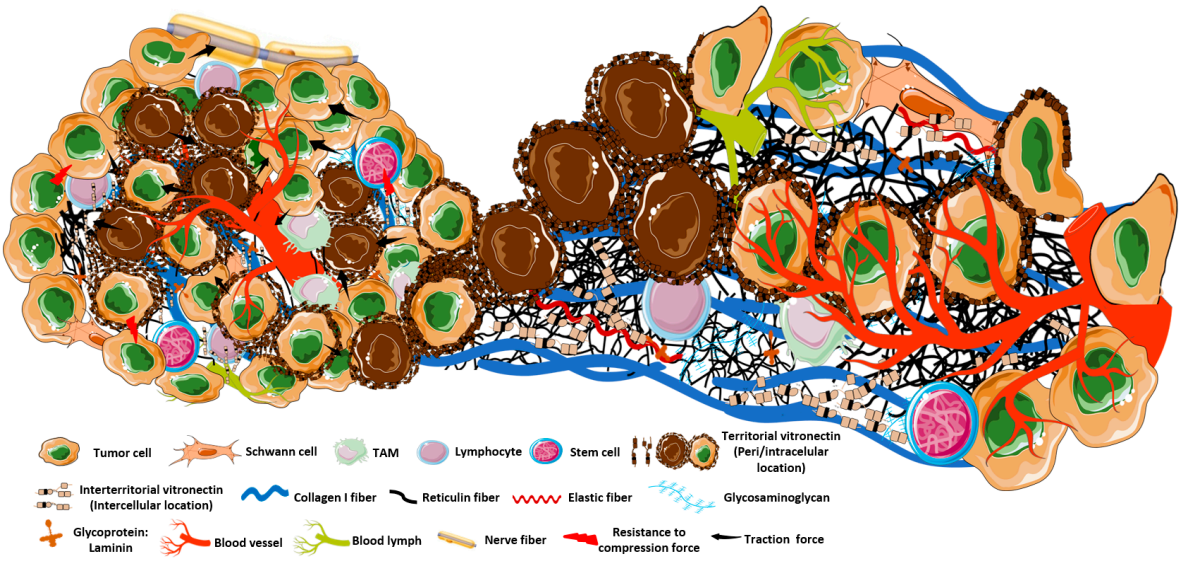

Figure 22. The tumour microenvironment on a high-risk scenario, where VN, hypothetically, creates tracks to aid the tumoral cells to invade other organs. On this figure, tumour cells are segregating VN (brown colour) while changing the rest of the TME elements to help them to survive and thrive on the organism it is living upon.

Overall, we have seen VN as the most significant marker of our dataset. Nonetheless, when coupling all the variables (VN and other TME elements), another statistically relevant characteristic appeared: the aspect ratio of blood vessels. In particular, we found that more irregular blood vessels would be correlated to poorer prognosis of the patients. Irregular blood vessels were already associated with more aggressive tumours (Nagy et al., 2009; De Palma et al., 2017). This effect would be produced by an irregular angiogenesis triggered by the tumoral cells whose supply requirements are larger than non-tumoral cells. This is consistent with previous studies of angiogenesis in tumoral environments (Nagy et al., 2009; De Palma et al., 2017).

In this thesis, we have unveiled VN as a new possible actor on NB tumorigenesis. Our results suggest that VN organization is statistically related to tumoral aggressiveness and the prognosis of NB patient. Interestingly, we found that VN would be the most influential factor compared to any other tested element of the TME in NB. Only a feature of blood vessels was found to be relevant for one of our criteria (Risk pre-treatment classification). Thus, none of the acquired features for collagen, reticulin, lymphatic vessels or GAGs came out as statistically significant, meaning that VN may play a greater role than previously assumed in NB. Although we need to deep down on the molecular basis of VN during NB, we highlight VN as a possible therapeutic target for NB. Further studies will be required to reveal the true role of VN in NB and to test whether our results are consistent.

6 Conclusions

1) We have analysed histopathological images from NB human patients, where different TME elements were displayed. These images where the initial source to apply the network approach.

2) We have developed a pipeline to fully analyse NB biopsies, combining topological and morphological features from VN and TME markers (collagen, reticulin, glycosaminoglycans, blood vessels and lymphatic vessels).

3) We, first, have focus on the role of VN in NB. For that purpose, we analysed two different types of VN: territorial (recently synthesized VN) and interterritorial VN (deposited in the ECM).

4) Considering only the VN results from the pure topological characteristics, we suggest that the way VN is distributed on the biopsy might be related to the prognosis of the patient.

5) Using the VN dataset (47 features), we performed a statistical pipeline consisted in two steps: a univariate analysis and a multivariate analysis. As a result, we obtained that two topological features from territorial VN were the most significant characteristics of the VN dataset: Euler number on risk pre-stratification group and Branching on genetic instability criterion.

6) To extend the same type of analysis to other TME elements, we collected features from collagen, reticulin, GAGs, blood vessels and lymphatics vessels to create the TME dataset (114 features).

7) By performing the same statistical analysis, we obtained another significant feature related to the TME: the aspect ratio of the blood vessels. Consistent with previous works, we found that irregular blood vessels were related to more aggressive tumours.

8) Altogether, from all the elements of the TME analysed, the most relevant marker in our dataset is VN. We hypothesize that VN would be creating migration tracks, changing dramatically the ECM that, in turn, will increase tumour aggressiveness.

7 References

Afratis, N., Gialeli, C., Nikitovic, D., Tsegenidis, T., Karousou, E., Theocharis, A.D., Pavão, M.S., Tzanakakis, G.N., and Karamanos, N.K. (2012). Glycosaminoglycans: key players in cancer cell biology and treatment. FEBS J. *279*, 1177–1197.

Ahmed, N.K., Neville, J., Rossi, R.A., Duffield, N.G., and Willke, T.L. (2017). Graphlet decomposition: framework, algorithms, and applications. Knowl. Inf. Syst. *50*, 689–722.

Albert, R., Jeong, H., and Barabási, A.-L. (1999). Diameter of the World-Wide Web. Nature *401*, 130–131.

Allison, P. (2004). Convergence Problems in Logistic Regression. pp. 238–252.

Almaas, E. (2007). Biological impacts and context of network theory. J. Exp. Biol. *210*, 1548–1558.

Angel Arul Jothi, J., and Mary Anita Rajam, V. (2017). A survey on automated cancer diagnosis from histopathology images. Artif. Intell. Rev. *48*, 31–81.

Barabási, A.L., and Oltvai, Z.N. (2004). Network biology: Understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Belsare, A.D., and Mushrif, M.M. (2012). Histopathological Image Analysis using Image Processing Techniques: An Overview. Int. J. Signal Image Process. *3*, 23–36.

Binchi, J., Merelli, E., Rucco, M., Petri, G., and Vaccarino, F. (2014). JHoles: A tool for understanding biological complex networks via clique weight rank persistent homology. Electron. Notes Theor. Comput. Sci. *306*, 5–18.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. Phys. Rep. *424*, 175–308.

Borriello, L., Seeger, R.C., Asgharzadeh, S., and DeClerck, Y.A. (2016). More than the genes, the tumor microenvironment in neuroblastoma. Cancer Lett. *380*, 304–314.

Burgos-Panadero, R., Noguera, I., Cañete, A., Navarro, S., and Noguera, R. (2019). Vitronectin as a molecular player of the tumor microenvironment in neuroblastoma. BMC Cancer *19*, 479.

Calcagno, V., and Mazancourt, C. de (2010). glmulti : An R Package for Easy Automated Model Selection with (Generalized) Linear Models. J. Stat. Softw. *34*, 1–29.

Cannoodt, R., Ruyssinck, J., Ramon, J., De Preter, K., and Saeys, Y. (2018). IncGraph: Incremental graphlet counting for topology optimisation. PLoS One *13*, e0195997.

Chung, F.F.R.K., and Graham, F. (1997). Spectral graph theory (American Mathematical Soc.).

Cohn, S.L., Pearson, A.D.J., London, W.B., Monclair, T., Ambros, P.F., Brodeur, G.M., Faldum, A., Hero, B., Iehara, T., Machin, D., et al. (2009). The International Neuroblastoma Risk Group (INRG) classification system: An INRG task force report. J. Clin. Oncol. *27*, 289–297.

DeClerck, Y.A., Mercurio, A.M., Stack, M.S., Chapman, H.A., Zutter, M.M., Muschel, R.J., Raz, A., Matrisian, L.M., Sloane, B.F., Noel, A., et al. (2004). Proteases, Extracellular Matrix, and Cancer: A Workshop of the Path B Study Section. Am. J. Pathol. *164*, 1131–1139.

Dekkers, M.P.J., Nikoletopoulou, V., and Barde, Y.A. (2013). Death of developing neurons: New insights and implications for connectivity. J. Cell Biol. *203*, 385–393.

Dennie, C., Thornhill, R., Sethi-Virmani, V., Souza, C.A., Bayanati, H., Gupta, A., and Maziak, D. (2016). Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. Quant. Imaging Med. Surg. *6*, 6–15.

Dimova, T., Georgieva, R., Petlov, D., and Lazarova, S. (2005). Comparative study on the expression of fibronectin, vitronectin, α5β1, β3 and αv integrins in goat and pig adrenal glands. Comptes Rendus l'Academie Bulg. Des Sci. *58*, 1457–1462.

Discher, D.E., Janmey, P., and Wang, Y.-L. (2005). Tissue cells feel and respond to the stiffness of their substrate. Science *310*, 1139–1143.

Dong, J., and Horvath, S. (2007). Understanding network concepts in modules. BMC Syst. Biol. *1*, 24.

Doyle, S., Agner, S., Madabhushi, A., Feldman, M., and Tomaszewski, J. (2008). Automated grading of breast cancer histopathology using spectral clusteringwith textural and architectural image features. In 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (IEEE), pp. 496–499.

Du, Q., Faber, V., and Gunzburger, M. (1999). Centroidal Voronoi Tessellations:Applications and Algorithms. SIAM Rev. *41*, 637–676.

Ediger, D., Jiang, K., Riedy, J., Bader, D.A., and Corley, C. (2010). Massive Social Network Analysis: Mining Twitter for Social Good. In 2010 39th International Conference on Parallel Processing, (IEEE), pp. 583–593.

Emon, B., Bauer, J., Jain, Y., Jung, B., and Saif, T. (2018). Biophysics of Tumor Microenvironment and Cancer Metastasis - A Mini Review. Comput. Struct. Biotechnol. J. *16*, 279–287.

Escolano, F., Hancock, E.R., and Lozano, M.A. (2012). Heat diffusion: Thermodynamic depth complexity of networks. Phys. Rev. E *85*, 036206.

Escudero, L.M., da F. Costa, L., Kicheva, A., Briscoe, J., Freeman, M., and Babu, M.M.M. (2011). Epithelial organisation revealed by a network of cellular contacts. Nat. Commun. *2*, 526.

Estrada, E. (2006). Protein bipartivity and essentiality in the yeast protein-protein interaction network. J. Proteome Res. *5*, 2177–2184.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein–protein interactions. Nature *340*, 245–246.

Fiorino, S., Bacchi-Reggiani, L., Pontoriero, L., Gallo, C., Chili, E., Masetti, M., Zanini, N., Grondona, A., Silvestri, T., Deleonardi, G., et al. (2014). Tensegrity model hypothesis: may this paradigm be useful to explain hepatic and pancreatic carcinogenesis in patients with persistent hepatitis B or hepatitis C virus infection? JOP *15*, 151–164.

Gilkes, D.M., Semenza, G.L., and Wirtz, D. (2014). Hypoxia and the extracellular matrix: Drivers of tumour metastasis. Nat. Rev. Cancer *14*, 430–439.

Gómez-Gálvez, P., Vicente-Munuera, P., Tagua, A., Forja, C., Castro, A.M.A.M., Letrán, M., Valencia-Expósito, A., Grima, C., Bermúdez-Gallardo, M., Serrano-Pérez-Higueras, Ó., et al. (2018). Scutoids are a geometrical solution to three-dimensional packing of epithelia. Nat. Commun. *9*, 2960.

Guirao, B., Rigaud, S.U., Bosveld, F., Bailles, A., López-Gay, J., Ishihara, S., Sugimura, K., Graner, F., and Bellaïche, Y. (2015). Unified quantitative characterization of epithelial tissue development. Elife *4*, 773–786.

Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., and Yener, B. (2009). Histopathological image analysis: a review. IEEE Rev. Biomed. Eng. *2*, 147–171.

Hansford, L.M., Thomas, W.D., Keating, J.M., Burkhart, C.A., Peaston, A.E., Norris, M.D., Haber, M., Armati, P.J., Weiss, W.A., and Marshall, G.M. (2004). Mechanisms of embryonal tumor initiation: Distinct roles for MycN expression and MYCN amplification. Proc. Natl. Acad. Sci. U. S. A. *101*, 12664–12669.

He, X., and Zhang, J. (2006). Why Do Hubs Tend to Be Essential in Protein Networks? PLoS Genet. *2*, e88.

Healy, K. (2005). Book Review: An R and S-PLUS Companion to Applied Regression. Sociol. Methods Res. *34*, 137–140.

Heinze, G., and Schemper, M. (2002). A solution to the problem of separation in logistic regression. Stat. Med. *21*, 2409–2419.

Heller, D., Hoppe, A., Restrepo, S., Gatti, L., Tournier, A.L.L., Tapon, N., Basler, K., and Mao, Y.

(2016). EpiTools: An Open-Source Image Analysis Toolkit for Quantifying Epithelial Growth Dynamics. *36*, 103–116.

Hočevar, T., and Demšar, J. (2014). A combinatorial approach to graphlet counting. Bioinformatics *30*, 559–565.

Holme, P., Kim, B.J., Yoon, C.N., and Han, S.K. (2002). Attack vulnerability of complex networks. Phys. Rev. E *65*, 056109.

Howarth, P., and Rüger, S. (2004). Evaluation of Texture Features for Content-Based Image Retrieval. In Proceedings of the International Conference on Image and Video Retrieval, pp. 326–334.

Huang, S., and Ingber, D.E. (2005). Cell tension, matrix mechanics, and cancer development. Cancer Cell *8*, 175–176.

Huang, C., Holfeld, J., Schaden, W., Orgill, D., and Ogawa, R. (2013). Mechanotherapy: revisiting physical therapy and recruiting mechanobiology for a new era in medicine. Trends Mol. Med. *19*, 555–564.

Ingber, D.E. (2002). Cancer as a disease of epithelial-mesenchymal interactions and extracellular matrix regulation. Differentiation *70*, 547–560.

Ingber, D., and Jamieson, J. (1985). Cells as tensegrity structures: Architectural regulation of histodifferentiationby physical forces tranduced over basement membranes. Gene Expr. Dur. Norm. Malig. Differ. 13–32.

Ingber, D.E., Madri, J.A., and Jamieson, J.D. (1981). Role of basal lamina in neoplastic disorganization of tissue architecture. Proc. Natl. Acad. Sci. U. S. A. *78*, 3901–3905.

Isogai, C., Laug, W.E., Shimada, H., Declerck, P.J., Stins, M.F., Durden, D.L., Erdreich-Epstein, A., and DeClerck, Y.A. (2001). Plasminogen activator inhibitor-1 promotes angiogenesis by stimulating endothelial cell migration toward fibronectin. Cancer Res. *61*, 5587–5594.

Jaffe, N. (1976). Neuroblastoma: Review of the literature and an examination of factors contributing to its enigmatic character. Cancer Treat. Rev. *3*, 61–82.

Jeong, H., Mason, S.P., Barabási, A.-L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41–42.

Junker, B.H., and Schreiber, F. (2008). Analysis of Biological Networks (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Kay, E., Bondy, J. a, and Murty, U.S.R. (1977). Graph Theory with Applications. Oper. Res. Q. *28*,

237.

Kenny, H.A., Kaur, S., Coussens, L.M., and Lengyel, E. (2008). The initial steps of ovarian cancer cell metastasis are mediated by MMP-2 cleavage of vitronectin and fibronectin. J. Clin. Invest. *118*, 1367–1379.

Kessenbrock, K., Wang, C.-Y., and Werb, Z. (2015). Matrix metalloproteinases in stem cell regulation and cancer. Matrix Biol. *44–46*, 184–190.

Kim, S.-H., Turnbull, J., and Guimond, S. (2011). Extracellular matrix and cell signalling: the dynamic cooperation of integrin, proteoglycan and growth factor receptor. J. Endocrinol. *209*, 139–151.

Kraniak, J.M., Mattingly, R.R., and Sloane, B.F. (2018). Roles of Pericellular Proteases in Tumor Angiogenesis : Therapeutic Implications. Extracell. Target. Cell Signal. Cancer 411–446.

Leavesley, D.I., Kashyap, A.S., Croll, T., Sivaramakrishnan, M., Shokoohmand, A., Hollier, B.G., and Upton, Z. (2013). Vitronectin--master controller or micromanager? IUBMB Life *65*, 807–818.

Li, H., Fan, X., and Houghton, J. (2007). Tumor microenvironment: The role of the tumor stroma in cancer. J. Cell. Biochem. *101*, 805–815.

Li, R., Ren, M., Chen, N., Luo, M., Zhang, Z., and Wu, J. (2012). Vitronectin increases vascular permeability by promoting VE-cadherin internalization at cell junctions. PLoS One *7*, e37195.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Trans. Inf. Theory *28*, 129–137.

Lo, C.-M., Wang, H.-B., Dembo, M., and Wang, Y. (2000). Cell Movement Is Guided by the Rigidity of the Substrate. Biophys. J. *79*, 144–152.

De Lorenzi, V., Sarra Ferraris, G.M., Madsen, J.B., Lupia, M., Andreasen, P.A., and Sidenius, N. (2016). Urokinase links plasminogen activation and cell adhesion by cleavage of the RGD motif in vitronectin. EMBO Rep. *17*, 982–998.

Madsen, C.D., and Sidenius, N. (2008). The interaction between urokinase receptor and vitronectin in cell adhesion and signalling. Eur. J. Cell Biol. *87*, 617–629.

Madsen, C.D., Ferraris, G.M.S., Andolfo, A., Cunningham, O., and Sidenius, N. (2007). uPAR-induced cell adhesion and migration: vitronectin provides the key. J. Cell Biol. *177*, 927–939.

Mäkinen, T., Norrmén, C., and Petrova, T. V. (2007). Molecular mechanisms of lymphatic vascular development. Cell. Mol. Life Sci. *64*, 1915–1929.

Mandriota, S.J., Jussila, L., Jeltsch, M., Compagni, A., Baetens, D., Prevo, R., Banerji, S., Huarte, J., Montesano, R., Jackson, D.G., et al. (2001). Vascular endothelial growth factor-C-mediated

lymphangiogenesis promotes tumour metastasis. EMBO J. *20*, 672–682.

Maris, J.M. (2010). Recent Advances in Neuroblastoma. N. Engl. J. Med. *362*, 2202–2211.

Marshall, G.M., Carter, D.R., Cheung, B.B., Liu, T., Mateos, M.K., Meyerowitz, J.G., and Weiss, W.A. (2014). The prenatal origins of cancer. Nat. Rev. Cancer *14*, 277–289.

Martín-Rodríguez, J.F., Madrazo-Atutxa, A., Venegas-Moreno, E., Benito-López, P., Gálvez, M.Á., Cano, D.A., Tinahones, F.J., Torres-Vela, E., Soto-Moreno, A., and Leal-Cerro, A. (2013). Neurocognitive Function in Acromegaly after Surgical Resection of GH-Secreting Adenoma versus Naïve Acromegaly. PLoS One *8*, e60041.

Martín-Rodríguez, J.F., Cervera-Barajas, A., Madrazo-Atutxa, A., García-Luna, P.P., Pereira, J.L., Castro-Luque, J., León-Justel, A., Morales-Conde, S., Castillo, J.R., Leal-Cerro, A., et al. (2014). Effect of bariatric surgery on microvascular dysfunction associated to metabolic syndrome: A 12-month prospective study. Int. J. Obes. *38*, 1410–1415.

McCormack, M.E., Lopez, J.A., Crocker, T.H., and Mukhtar, M.S. (2016). Making the right connections: Network biology and plant immune system dynamics. Curr. Plant Biol. *5*, 2–12.

McLeod, A.I., and Xu, C. (2017). bestglm: Best Subset GLM and Regression Utilities.

Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement - IMC '07, (New York, New York, USA: ACM Press), p. 29.

Nagy, J.A., Chang, S.-H., Dvorak, A.M., and Dvorak, H.F. (2009). Why are tumour blood vessels abnormal and why is it important to know? Br. J. Cancer *100*, 865–869.

Newman, M.E.J. (2003). The Structure and Function of Complex Networks. SIAM Rev. *45*, 167–256.

Ng, M.R., and Brugge, J.S. (2009). A Stiff Blow from the Stroma: Collagen Crosslinking Drives Tumor Progression. Cancer Cell *16*, 455–457.

Øien, A.H., and Wiig, H. (2016). Electrostatic, elastic and hydration-dependent interactions in dermis influencing volume exclusion and macromolecular transport. J. Theor. Biol. *400*, 80–91.

Orr, F.W., Podor, T.J., Buchanan, M.R., and Orr, F.W. (1992). Up-regulated Biosynthesis and Expression of Endothelial Cell Vitronectin Receptor Enhances Cancer Cell Adhesion. Cancer Res. *52*, 2202–2208.

De Palma, M., Biziato, D., and Petrova, T. V. (2017). Microenvironmental regulation of tumour angiogenesis. Nat. Rev. Cancer *17*, 457–474.

Pathak, A.P., Artemov, D., Neeman, M., and Bhujwalla, Z.M. (2006). Lymph Node Metastasis in Breast Cancer Xenografts Is Associated with Increased Regions of Extravascular Drain, Lymphatic Vessel Area, and Invasive Phenotype. Cancer Res *66*, 5151–5159.

Petrie, A.G. (2015). Introduction to Regression and Modeling with R (Cognella Academic Publishing).

Pickup, M.W., Mouw, J.K., and Weaver, V.M. (2014). The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. *15*, 1243–1253.

Pijuan-Thompson, V., and Gladson, C.L. (1997). Ligation of integrin alpha5beta1 is required for internalization of vitronectin by integrin alphavbeta3. J. Biol. Chem. *272*, 2736–2743.

Pinto, N.R., Applebaum, M.A., Volchenboum, S.L., Matthay, K.K., London, W.B., Ambros, P.F., Nakagawara, A., Berthold, F., Schleiermacher, G., Park, J.R., et al. (2015). Advances in Risk Classification and Treatment Strategies for Neuroblastoma. J. Clin. Oncol. *33*, 3008–3017.

Ploner, M., Dunkler, D., Southworth, H., and Heinz, G. (2010). logistf: Firth's bias reduction logistic regression. R Packag. Version 1.10.

Preissner, K.T. (1989). The role of vitronectin as multifunctional regulator in the hemostatic and immune systems. Blut *59*, 419–431.

Preissner, K.T. (1991). Structure and Biological Role of Vitronectin. Annu. Rev. Cell Biol. *7*, 275–310.

Preissner, K., and Reuning, U. (2011). Vitronectin in Vascular Context: Facets of a Multitalented Matricellular Protein. Semin. Thromb. Hemost. *37*, 408–424.

Przulj, N. (2007). Biological network comparison using graphlet degree distribution. Bioinformatics *23*, e177–e183.

Pržulj, N., Corneil, D.G., and Jurisica, I. (2004). Modeling interactome: Scale-free or geometric? Bioinformatics *20*, 3508–3515.

La Quaglia, M.P. (2014). State of the art in oncology: High risk neuroblastoma, alveolar rhabdomyosarcoma, desmoplastic small round cell tumor, and POST-TEXT 3 and 4 hepatoblastoma. J. Pediatr. Surg. *49*, 233–240.

Rojo, M.G., Bueno, G., and Slodkowska, J. (2010). Review of imaging solutions for integrated quantitative immunohistochemistry in the Pathology daily practice. Folia Histochem. Cytobiol. *47*, 349–354.

Sáez, A., Rivas, E., Montero-Sánchez, A., Paradas, C., Acha, B., Pascual, A., Serrano, C., and

Escudero, L.M. (2013). Quantifiable diagnosis of muscular dystrophies and neurogenic atrophies through network analysis. BMC Med. *11*, 77.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1987). Akaike information criterion statistics. Math. Comput. Simul. *29*, 452.

Sanchez-Gutierrez, D., Saez, A., Pascual, A., and Escudero, L.M. (2013). Topological progression in proliferating epithelia is driven by a unique variation in polygon distribution. PLoS One *8*, e79227.

Scott, J. (1988). Social Network Analysis. Sociology *22*, 109–127.

Seiffert, D. (1997). Constitutive and regulated expression of vitronectin. Histol Histopathol *12*, 787–797.

Sevenich, L., and Joyce, J.A. (2014). Pericellular proteolysis in cancer. Genes Dev. *28*, 2331–2347.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. *13*, 2498–2504.

Stamenović, D., and Ingber, D.E. (2009). Tensegrity-guided self assembly: from molecules to living cells. Soft Matter *5*, 1137–1145.

Strogatz, S.H. (2001). Exploring complex networks. Nature *410*, 268–276.

Tadeo, I., Berbegall, A.P., Escudero, L.M., Alvaro, T., and Noguera, R. (2014). Biotensegrity of the Extracellular Matrix: Physiology, Dynamic Mechanical Balance, and Implications in Oncology and Mechanotherapy. Front. Oncol. *4*, 39.

Tadeo, I., Berbegall, A.P., Castel, V., García-Miguel, P., Callaghan, R., Påhlman, S., Navarro, S., and Noguera, R. (2016). Extracellular matrix composition defines an ultra-high-risk group of neuroblastoma within the high-risk patient cohort. Br. J. Cancer *115*, 480–489.

Tadeo, I., Berbegall, A.P., Navarro, S., Castel, V., and Noguera, R. (2017). A stiff extracellular matrix is associated with malignancy in peripheral neuroblastic tumors. Pediatr. Blood Cancer *64*, e26449.

Tadeo, I., Gamero-Sandemetrio, E., Berbegall, A.P., Gironella, M., Ritort, F., Cañete, A., Bueno, G., Navarro, S., and Noguera, R. (2018). Lymph microvascularization as a prognostic indicator in neuroblastoma. Oncotarget *9*, 26157–26170.

Tarr, G., Müller, S., and Welsh, A.H. (2018). mplot : An R Package for Graphical Model Stability and Variable Selection Procedures. J. Stat. Softw. *83*.

Team, R.D.C., and R Development Core Team, R. (2016). R: A Language and Environment for

Statistical Computing. R Found. Stat. Comput.

Titford, M. (2006). A Short History of Histopathology Technique. J. Histotechnol. *29*, 99–110.

Toromanov, G., Gugutkov, D., Gustavsson, J., Planell, J., Salmerón-Sánchez, M., and Altankov, G. (2015). Dynamic Behavior of Vitronectin at the Cell-Material Interface. ACS Biomater. Sci. Eng. *1*, 927–934.

Tsuboi, A., Ohsawa, S., Umetsu, D., Sando, Y., Kuranaga, E., Igaki, T., and Fujimoto, K. (2018). Competition for Space Is Controlled by Apoptosis-Induced Change of Local Epithelial Topology. Curr. Biol. *28*, 2115–2128.

Tuttle, T.M. (2004). Technical advances in sentinel lymph node biopsy for breast cancer. Am. Surg. *70*, 407–413.

Valverde, S., Ohse, S., Turalska, M., West, B.J., and Garcia-Ojalvo, J. (2015). Structural determinants of criticality in biological networks. Front. Physiol. *6*, 127.

Veta, M., Pluim, J.P.W., Van Diest, P.J., and Viergever, M.A. (2014). Breast cancer histopathology image analysis: A review. IEEE Trans. Biomed. Eng. *61*, 1400–1411.

Vicente-Munuera, P., Gómez-Gálvez, P., Tetley, R.J., Forja, C., Tagua, A., Letrán, M., Tozluoglu, M., Mao, Y., and Escudero, L.M. (2020). EpiGraph: an open-source platform to quantify epithelial organization. Bioinformatics *36*, 1314–1316.

Vicente-Munuera, P., Burgos-Panadero, R., Noguera, I., Navarro, S., Noguera, R., Escudero, L.M.L.M., Vicente-Munuera, P., Burgos-Panadero, R., Noguera, I., Navarro, S., et al. (2020). The topology of vitronectin: A complementary feature for neuroblastoma risk classification based on computer-aided detection. Int. J. Cancer *146*, 553–565.

Viola, J.M., Porter, C.M., Gupta, A., Alibekova, M., Prahl, L.S., and Hughes, A.J. (2019). Kinomorphs: Shape-shifting tissues for developmental engineering. BioRxiv 768218.

Virchow, R. (1865). Die krankhaften Geschwülste.

Voûte, P.A., de Kraker, J., and Hoefnagel, C.A. (1992). Tumours of the Sympathetic Nervous System: Neuroblastoma, Ganglioneuroma and Phaeochromocytoma. In Cancer in Children, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 226–243.

Wittke, C., Mayer, J., and Schweiggert, F. (2007). On the classification of prostate carcinoma with methods from spatial statistics. IEEE Trans. Inf. Technol. Biomed. *11*, 406–414.

Wright, J.H. (1910). Neurocytoma or Neuroblastoma, a kind of tumor not generally recognized. J. Exp. Med. *12*, 556–561.

Yao, J., Ganti, D., Luo, X., Xiao, G., Xie, Y., Yan, S., and Huang, J. (2015). Computer-assisted diagnosis of lung cancer using quantitative topology features. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), (Springer, Cham), pp. 288–295.

Yaveroğlu, Ö.N., Malod-Dognin, N., Davis, D., Levnajic, Z., Janjic, V., Karapandza, R., Stojmirovic, A., and Pržulj, N. (2014). Revealing the Hidden Language of Complex Networks. Sci. Rep. *4*, 1–9.

Yu, H., Kim, P.M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. PLoS Comput. Biol. *3*, 713–720.

Zhang, H., Qi, M., Li, S., Qi, T., Mei, H., Huang, K., Zheng, L., and Tong, Q. (2012). microRNA-9 Targets Matrix Metalloproteinase 14 to Inhibit Invasion, Metastasis, and Angiogenesis of Neuroblastoma Cells. Mol. Cancer Ther. *11*, 1454–1466.

Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.M., Witte, O.N., Toran-Allerand, D., Gee, C.E., et al. (1986). Differential expression of myc family genes during murine development. Nature *319*, 780–783.

Topological

ID	Name
48	Reticulin - Sorting tensegral index
49	Reticulin - Iteration tensegral index
50	Reticulin - MST tensegral index
51	Reticulin - std percentage stained area per region
52	Reticulin - std percentage stained area per node
53	Reticulin - mean quantity of branches per region
54	Reticulin - mean quantity of branches per node
55	Reticulin - euler number per stained area
56	Reticulin - euler number per region
57	Reticulin - euler number per node
58	Reticulin - number of holes per stained area
59	Reticulin - std area of holes
74	Blood vessels - Sorting tensegral index
75	Blood vessels - Iteration tensegral index
76	Blood vessels - MST tensegral index
77	Blood vessels - std percentage stained area per region
78	Blood vessels - std percentage stained area per node
79	Blood vessels - mean quantity of branches per region
80	Blood vessels - mean quantity of branches per node
81	Blood vessels - euler number per stained area
82	Blood vessels - euler number per region
83	Blood vessels - euler number per node
84	Blood vessels - number of holes per stained area
85	Blood vessels - std area of holes
102	Collagen - Sorting tensegral index
103	Collagen - Iteration tensegral index
104	Collagen - MST tensegral index
105	Collagen - std percentage stained area per region
106	Collagen - std percentage stained area per node
107	Collagen - mean quantity of branches per region
108	Collagen - mean quantity of branches per node
109	Collagen - euler number per stained area
110	Collagen - euler number per region
111	Collagen - euler number per node
112	Collagen - number of holes per stained area
113	Collagen - std area of holes
118	GAGs - Sorting tensegral index
119	GAGs - Iteration tensegral index
120	GAGs - MST tensegral index
121	GAGs - std percentage stained area per region
122	GAGs - std percentage stained area per node
123	GAGs - mean quantity of branches per region
124	GAGs - mean quantity of branches per node
125	GAGs - euler number per stained area
126	GAGs - euler number per region
127	GAGs - euler number per node
128	GAGs - number of holes per stained area
129	GAGs - std area of holes
134	Lymphatic vessels - Sorting tensegral index
135	Lymphatic vessels - Iteration tensegral index
136	Lymphatic vessels - MST tensegral index
137	Lymphatic vessels - std percentage stained area per region
138	Lymphatic vessels - std percentage stained area per node
139	Lymphatic vessels - mean quantity of branches per region
140	Lymphatic vessels - mean quantity of branches per node
141	Lymphatic vessels - euler number per stained area
142	Lymphatic vessels - euler number per region
143	Lymphatic vessels - euler number per node
144	Lymphatic vessels - number of holes per stained area
145	Lymphatic vessels - std area of holes

Non-Topological

ID	Name
60	Reticulin - mean percentage stained area per region
61	Reticulin - mean percentage stained area per node
62	Reticulin - mean area of holes
63	Reticulin - Stained area / mm2
64	Reticulin - Stained area
65	Reticulin - Mean area
66	Reticulin - Aspect
67	Reticulin - Roundness
68	Reticulin - Length
69	Reticulin - Width
70	Reticulin - Per_ratio
71	Reticulin - F_dimen
72	Reticulin - Dendrites
73	Reticulin - Angle
86	Blood vessels - mean percentage stained area per region
87	Blood vessels - mean percentage stained area per node
88	Blood vessels - mean area of holes
89	Blood vessels - Mean area
90	Blood vessels - Width
91	Blood vessels - Length
92	Blood vessels - Angle
93	Blood vessels - Roundness
94	Blood vessels - Aspect
95	Blood vessels - PerRatio
96	Blood vessels - Perimeter
97	Blood vessels - Deformity
98	Blood vessels - Shape
99	Blood vessels - Vertices
100	Blood vessels - Vessels/mm2
101	Blood vessels - area_clasearea_cilindro
114	Collagen - mean percentage stained area per region
115	Collagen - mean percentage stained area per node
116	Collagen - mean area of holes
117	Collagen - Stained area
130	GAGs - mean percentage stained area per region
131	GAGs - mean percentage stained area per node
132	GAGs - mean area of holes
133	GAGs - Stained area
146	Lymphatic vessels - mean percentage stained area per region
147	Lymphatic vessels - mean percentage stained area per node
148	Lymphatic vessels - mean area of holes
149	Lymphatic vessels - Mean area
150	Lymphatic vessels - Width
151	Lymphatic vessels - Length
152	Lymphatic vessels - Angle
153	Lymphatic vessels - Roundness
154	Lymphatic vessels - Aspect
155	Lymphatic vessels - PerRatio
156	Lymphatic vessels - Perimeter
157	Lymphatic vessels - Deformity
158	Lymphatic vessels - Shape
159	Lymphatic vessels - Vertices
160	Lymphatic vessels - Vessels/mm2
161	Lymphatic vessels - area_clasearea_cilindro

Additional Table 1. TME dataset. Additional dataset with the characteristics extracted from reticulin, blood vessels, collagen, GAGs and lymphatic vessels (**Material and Methods**). This table compliments **Table 2**.

“Never confuse education with intelligence, you can have a PhD and still be an idiot…”