

Pan tissue analysis of human *NUMB* alternative splicing.

Yangjing Zhang^{a,b} and C. Jane McGlade^{a,b,c*}

^a Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, ON, Canada, M5G 2M9

^b The Arthur and Sonia Labatt Brain Tumour Research Centre, ^c Program in Cell Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada, M5G 1X8

*Corresponding author: jmcglade@sickkids.ca, 416-813-8657

Keywords: Numb, protein isoforms, alternative splicing, exon 9, exon 3

Abstract

Alternative splicing (AS) of pre-mRNA generates multiple protein isoforms from a single gene. It is developmentally regulated and mis-regulation is associated with many diseases. Previous studies that examined AS of the cell fate determinant *NUMB* demonstrated that it undergoes a regulated switch in the inclusion of protein coding exons 3 and 9 during rodent development and the differentiation of human cell lines. Here we extend this work by comparing exon 9 and exon 3 inclusion levels of the human *NUMB* gene across different normal human tissues by analysis of RNA sequencing data deposited in the Genotype-Tissue Expression (GTEx) and the Vertebrate Alternative Splicing and Transcription (VastDB) databases. Our results support earlier studies and reveal specific Numb isoform expression patterns in previously unexamined tissue and cell types, suggesting that Numb isoform expression is regulated in the normal development of a broad range of tissue types throughout the body.

Introduction

Pre-mRNA undergoes splicing to remove introns and ligate exons to form mature mRNA. A single gene can generate multiple mRNA and protein isoforms by using alternative exons and splice sites, a process called alternative splicing (AS). AS occurs in nearly all multi-exon genes and diversifies the transcriptome and proteome. The protein isoforms produced by AS can have different, or even opposite, biological functions and deregulated AS often plays major roles in disease including cancer (Park, Pan et al. 2018). The *NUMB* gene product is a cell fate determinant that undergoes AS of two coding cassette exons (exon 3 and exon 9) to produce 4 mRNA transcripts resulting in 4 protein isoforms: p72 (Ex9in, Ex3in), p71 (Ex9in, Ex3sk), p66 (Ex9sk, Ex3in) and p65 (Ex9sk, Ex3sk) (Dho, French et al. 1999) (Figure 1 describes alternative splicing of *NUMB* RNA and formation of the major *NUMB* protein isoforms).

Individual studies have documented the switch in splicing of *NUMB* exon 9 from inclusion to exclusion during development in rodents and differentiation of cell line models (Dho, French et al. 1999, Verdi, Bashirullah et al. 1999, Dooley, James et al. 2003, Corallini, Fera et al. 2006, Bani-Yaghoub, Kubu et al. 2007, Gao, Chi et al. 2011, Moran, Goldberg et al. 2011, Kim, Nam et al. 2013). Furthermore, reversal of this switch to re-express exon 9 is a frequent AS event in cancer (Chen, Chen et al. 2009, Langer, Sohler et al. 2010, Misquitta-Ali, Cheng et al. 2011). Comprehensive analysis of exon 9 and exon 3 inclusion levels of the human *NUMB* gene utilizing the accumulating RNA-seq data deposited in public databases has not been reported.

The Genotype-Tissue Expression (GTEx) database is a comprehensive public resource to study tissue-specific gene expression and regulation in normal human samples from approximately 960 donors. The Cancer Genome Atlas (TCGA) provides an opportunity for the in-depth analysis of the transcriptome and AS in tumors from over 8,700 patients with matched normal samples spanning 33 cancer types (Kahles, Lehmann et al. 2018). Our recent study of *NUMB* AS using these datasets (Zhang, Dho et al. 2022) confirmed and extended earlier work showing increased exon 9 inclusion in multiple cancer types. Here, we describe *NUMB* splicing of exon 3 and exon 9 across different normal human tissues, derived from the analysis of the GTEx RNA-seq dataset. In addition, we compare the results with that from VastDB, a database of AS profiles across multiple tissue and cell types with a more limited sample size (Tapial, Ha et al. 2017). Our analysis not only confirms, using a much bigger sample size, previous studies, but also reports additional human tissue types that undergo AS regulation in *NUMB*.

Methods

Data download

Alternative splicing analysis was conducted across RNA sequencing data from the GTEx database (Kahles, Lehmann et al. 2018). The PSI (Percent Spliced In index) values for exon 9 (chr14:73745988 - 73746132) and exon 3 (chr14:73783097 - 73783130) regions for *NUMB* and the raw read counts of all aligned genes was extracted from the supplementary tables together with sample IDs. The PSI values for exon 9 (HsaEX0044216) and exon 3 (HsaEX0044220) for *NUMB* and the cFPKM (reads per kilobase of target transcript sequence per million of total

reads) of Numb and Numblike across different human tissues was downloaded from VastDB (http://vastdb.crg.eu/wiki/Main_Page) (Tapial, Ha et al. 2017).

Expression analysis

GTEX samples were categorized into different tissue types using sample metadata supplied on the GTEX websites. A dotplot representing the mean PSI value for samples of each tissue type was generated for exon 9 and exon 3. The DESeq2 (version 1.24.0) R software package was used to normalize the readcounts for aligned genes from the GTEX datasets. A dotplot representing the mean normalized read counts was generated for each normal tissue sample from the GTEX database using R (version 3.6.3). Dotplots for PSI value of NUMB exon 9 and exon 3, and the cFPKM (reads per kilobase of target transcript sequence per million of total reads) value of Numb and Numblike extracted from VastDB, were also generated using R.

Results and Discussion

The Percent Spliced In index (PSI) represents the percentage of gene specific mRNA transcripts which include a specific exon (Park, Pan et al. 2018)¹. We extracted the PSI value from the GTEX database for two cassette exons of the *NUMB* gene (coding exons 9 and exon 3) in different human tissues. There was an average of 52 samples per tissue type (Figure 1A, 1B, Supplementary Table 1) and the sample number, median/mean PSI value for *NUMB* exon 9 and exon 3, as well as total transcript level (in read counts) for *NUMB* are tabulated in Supplementary Table 1. Similarly, we extracted PSI values for *NUMB* exon 9 and exon 3 from the VastDB database, which includes AS profiles across multiple tissue and cell types but with a limited sample size (Supplementary Table 1) (~2 samples per cell or tissue type) (Figure 1C,

1D). Together these data provide a comprehensive profile of *NUMB* isoform expression across 83 different human tissues or cell types at the transcript level.

A broad range in exon 3 and exon 9 inclusion level is apparent across different cell and tissue types. Tissues were grouped into different categories based on patterns of exon 9 and exon 3 inclusion level (Table 1). In human stem cells (ESC: embryonic stem cell and iPS: induced pluripotent stem cells), both exon 9 and exon 3 inclusion is greater than 50% suggesting that the p72 (Ex9in & Ex3in) Numb isoform is the predominantly expressed Numb isoform (Figure 1C).

In adult brain and pituitary gland, immune and hematopoietic system, on average both exon 9 and exon 3 inclusion level is lower than 25%, suggesting that the p65 (Ex9sk & Ex3sk) Numb isoform is predominantly expressed (Table 1). These observations are consistent with previously observed changes in both exon 9 and exon 3 inclusion during development measured using RT-PCR or protein expression: rat and human brain (Verdi, Bashirullah et al. 1999, Dooley, James et al. 2003, Bani-Yaghoub, Kubu et al. 2007), mouse pituitary gland (Moran, Goldberg et al. 2011), rat auditory epithelium (Gao, Chi et al. 2011), and human and mouse erythroid cells (Huang, Vu et al. 2021).

Heart, adipose, breast, lung, adrenal, and gynecologic tissue (ovary, uterus, and fallopian tube) have nearly 0% exon 9 PSI and 25-50% exon 3 PSI on average, and therefore are predicted to express the exon 9 excluded p66 (Ex3in) and p65 (Ex3sk) isoforms predominantly.

In tissues with self-renewal activity and stem cell populations such as the testis and gastrointestinal tissue (stomach, small intestine and colon, pancreas), the exon 3-excluded *NUMB* isoforms, p71 (Ex9in) and p65 (Ex9sk) are predicted to be predominantly expressed since they exhibit ~50% exon 9 inclusion and less than 25% exon 3 inclusion. This is consistent

with published studies which examined both exon 9 and exon 3 inclusion in the development of mouse pancreas (Yoshida, Tokunaga et al. 2003) and testis (Corallini, Fera et al. 2006). Other tissues such as salivary gland, esophagus mucosa, kidney cortex and prostate, have similar levels of included and excluded exon 9 and exon 3 transcripts therefore potentially express similar levels of all four protein isoforms.

The total transcript level of *NUMB* and its parologue *NUMBL* were also compared across different human tissue types (Figure 2D). The mRNA transcript expression of *NUMB* appears to be lower in brain than in other tissues (Figure 2B). *NUMBL*, which exhibits some partially overlapping functions with *NUMB*, particularly in the nervous system (Petersen, Zou et al. 2002, Rasin, Gazula et al. 2007), is predominantly expressed in neural tissue. In most tissues, *NUMBL* is expressed at a lower level than *NUMB*, which is consistent with the literature (Figure 2D) (Zhong, Jiang et al. 1997, Gao, Chi et al. 2011).

This analysis reports the tissue specific expression of *NUMB* exon 9 and exon 3 in human tissues and confirms isoform expression patterns previously identified in tissues from mouse and cell line studies. Importantly, our study also reveals the specific Numb isoform expression patterns in previously unexamined tissue and cell types including the heart, gynecologic tissues, genitourinary tissues, immune and hematopoietic system, suggesting that the regulation of Numb isoform expression is likely critical in a broad range of tissue types throughout the body and reflects different roles for the isoforms in development (Figure 2).

Table 1 Numb isoform expression in normal human tissue

Tissue	Exon 9 PSI (%)	Exon 3 PSI (%)	Relative isoform abundance ¹	Source ²
Stem cell ESC iPS	50-75%	50-75%	p72 p71 p66 p65	VastDB
Adult brain tissue, pituitary tissue	0%	<25%	p66 p65	Both
Bone marrow, spleen, thymus, lymph node, WBC	<25%	<25%	p72 p71 p66 p65	VastDB
Adult heart, artery, muscle skeletal, adipose, breast, lung, adrenal, endothelium, fibroblasts, melanocytes, ovary, uterus, fallopian tube	0%	25-50%	p66 p65	Both
Skin sun exposed & not exposed, bladder, prostate, liver, thyroid, vagina, cervix, placenta	25%	50%	p72 p71 p66 p65	Both
Testis, stomach, small intestine, colon, pancreas	30-60%	<25%	p72 p71 p66 p65	Both
Minor salivary gland, esophagus mucosa, kidney cortex	50%	50%	p72 p71 p66 p65	Both

¹ Larger font represents greater relative abundance inferred from exon 9 and exon 3 inclusion.

² The range of exon 9 and exon 3 PSI was determined from values extracted from either the VastDB database or from both GTEx analysis and the VastDB database as indicated.

Figure 1

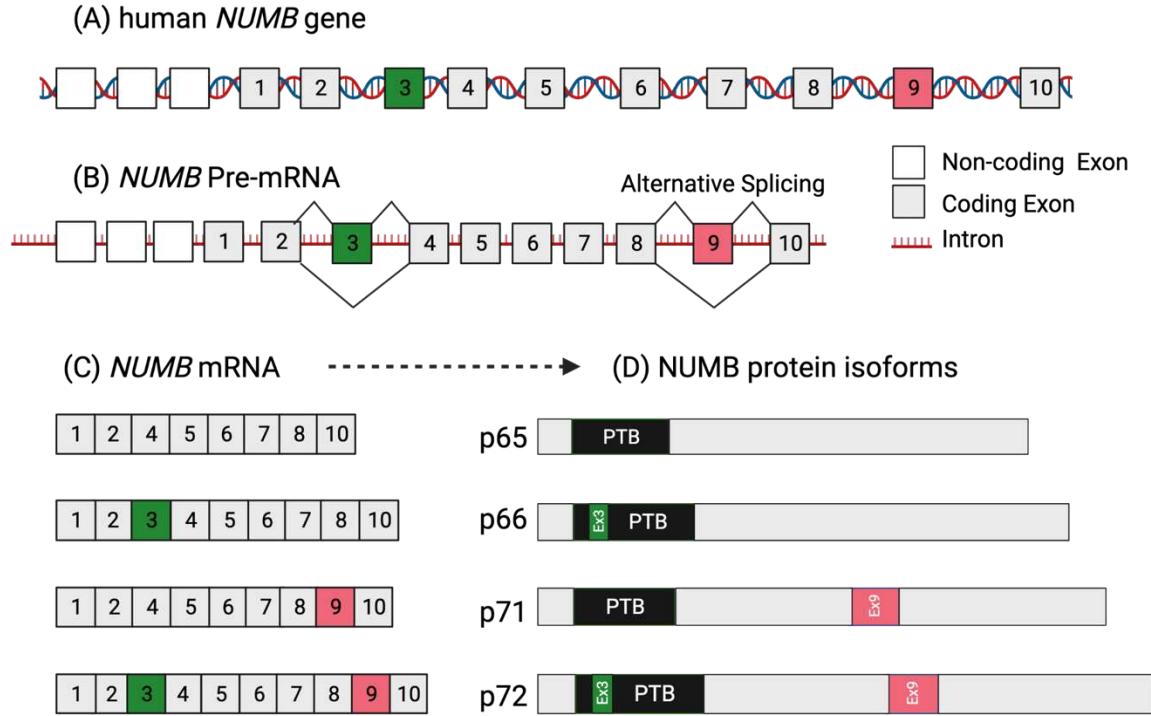
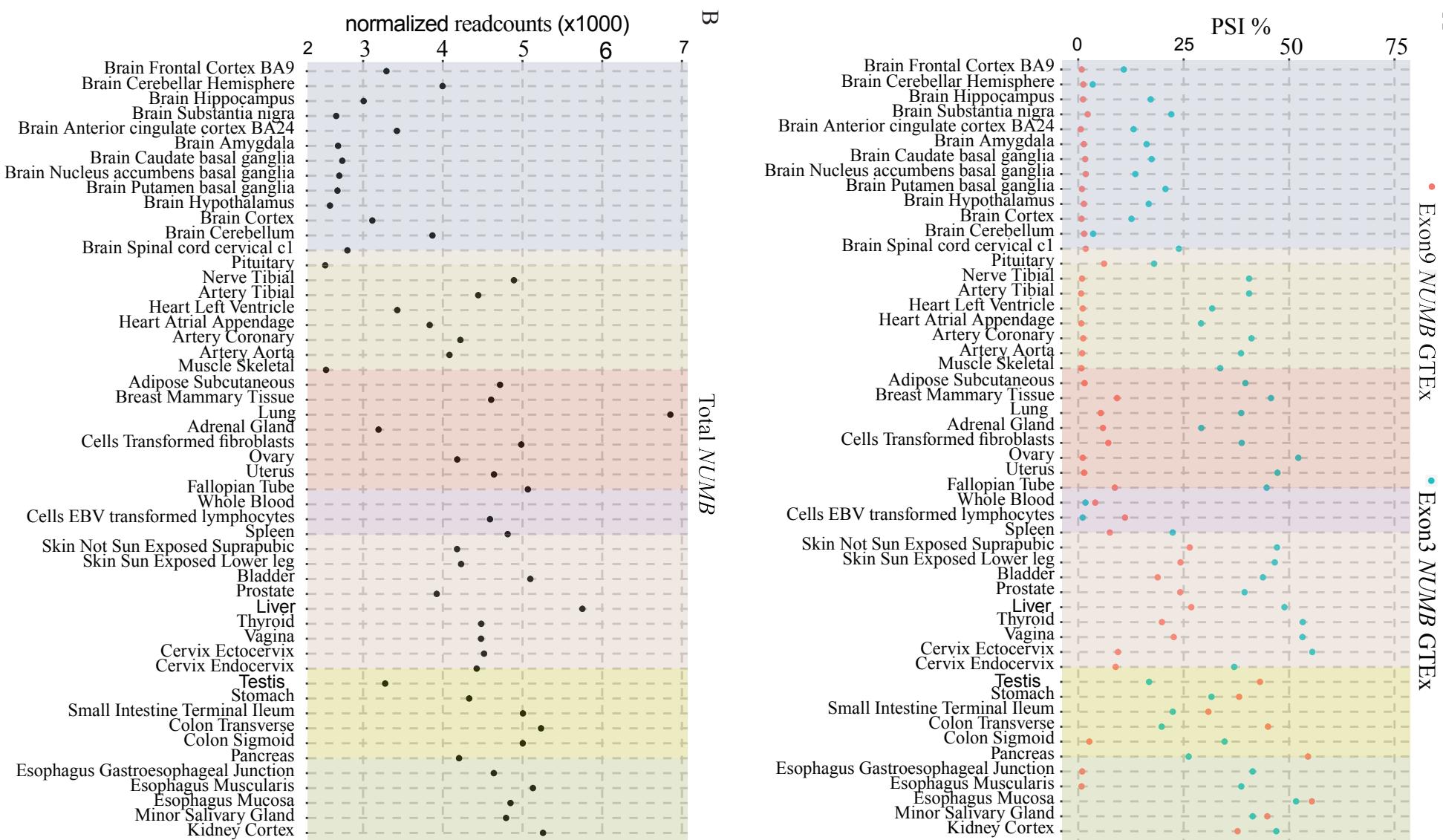


Figure 1 Formation of major *NUMB* protein isoforms by alternative splicing.

(A) The human *NUMB* gene is comprised of 13 exons: 1-3 are non-coding and 4 - 13 are coding (ENSEMBLE transcript: ENST00000555238.6). Exon numbering used here is based on the coding exons. (B and C) Alternative splicing of exons in precursor mRNA forms the four major *NUMB* mRNA splice variants by inclusion or exclusion of coding exon 3 (Ex3in and Ex3sk) and coding exon 9 (Ex9in and Ex9sk), and these are translated to *NUMB* protein isoforms identified here by their molecular weights: p65, p66, p71 and p72 (D).

Figure 2 bioRxiv preprint doi: <https://doi.org/10.1101/2023.09.26.559629>; this version posted September 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



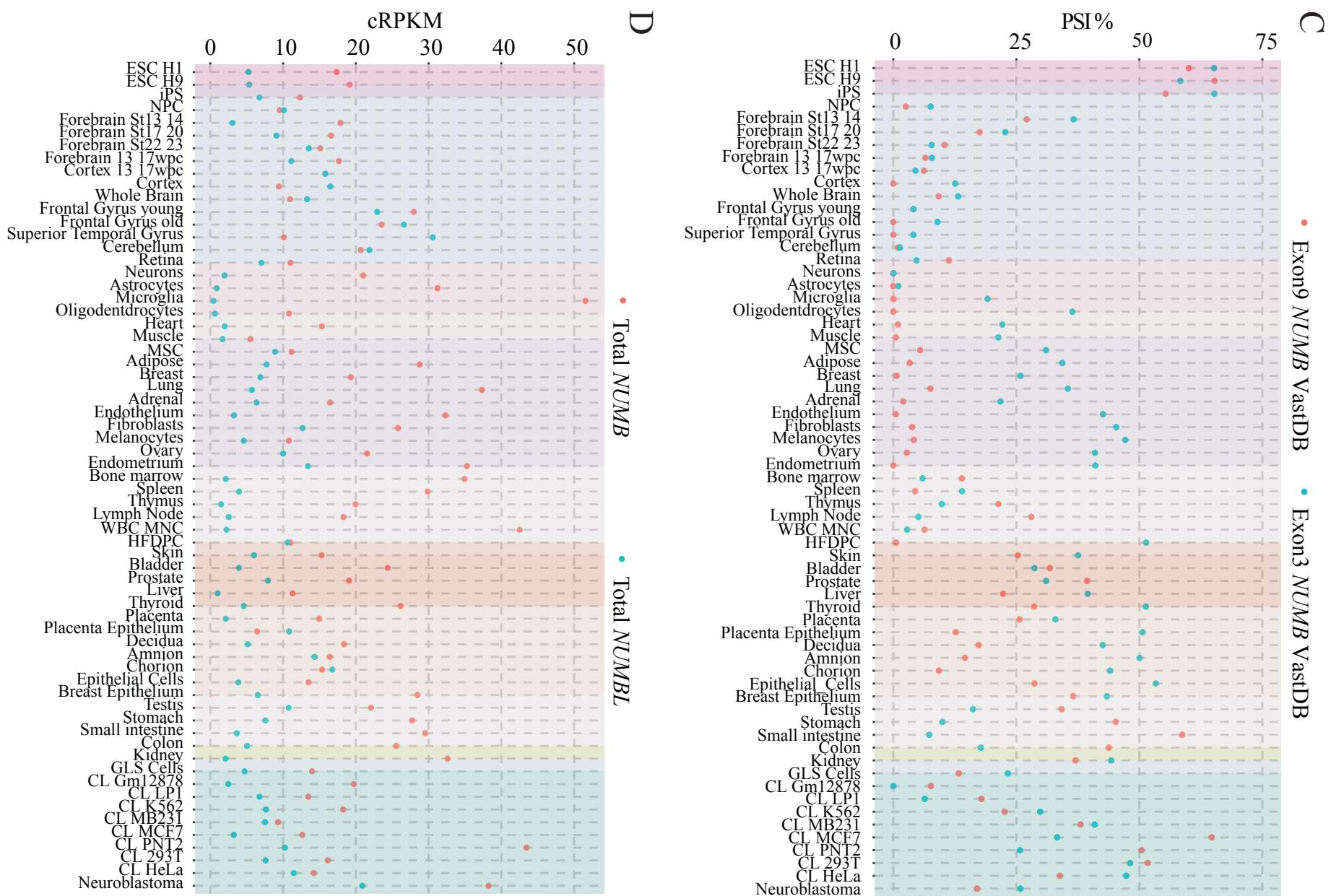


Figure 2 Mean inclusion level of exon 9 or exon 3 and total expression of *NUMB* (and *NUMBL*) in human tissues

(A & C) Exon 9 and exon 3 PSI in normal human tissues from the GTEx dataset (average sample size/tissue type = 52.3) (A) and in human tissues or cell lines from the VastDB dataset (average sample size/tissue type = 1.7) (C). Each dot represents the average PSI of samples for each tissue type. The represented tissues range from stem cells (ESC - embryonic stem cells, iPS - induced pluripotent stem cells etc.) to developing and adult neural tissues (different brain tissues etc.), glial cells (astrocytes, microglia, oligodendrocytes), muscle (heart, muscle), gynecologic tissue (ovary, uterus, and fallopian tube), genitourinary tissue (bladder, prostate, testis, kidney cortex), immune/hematopoiesis (bone marrow, spleen, thymus, lymph node, WBC), gastrointestinal tissues (stomach, small intestine and colon, pancreas) and other tissues. The rightmost columns for the VastDB dataset (C) are representative cell lines (CL). Primary neural progenitors, NPC. Hair follicle dermal papilla cells, HFDPC. Mesenchymal stem cells of adipose, MSC. Breast cancer cell line, MB231, MCF7. Prostatic cell line, PNT2. Lymphoblastoid cell line, Gm12878. Myeloma cell line, LP1. Chronic myelogenous leukemia cell line, K562.

(B) Total transcript expression of Numb in different human tissues from the GTEx dataset represented as normalized read counts.

(D) Total transcript expression of *NUMB* and *NUMBL* in different human tissue and cell lines from the VastDB dataset represented as cRPKM.

Acknowledgements: The authors thank Dr. Sascha Dho for comments on the manuscript. This work was supported with funds from the Canadian Institutes of Health Research to CJM (FRN 106507).

Author Contributions: YZ acquired, analyzed and interpreted data, drafted and critically reviewed the manuscript; CJM supervised YZ, analyzed and interpreted data, drafted and revised the manuscript.

Reference

Bani-Yaghoub, M., C. J. Kubu, R. Cowling, J. Rochira, G. N. Nikopoulos, S. Bellum and J. M. Verdi (2007). "A switch in numb isoforms is a critical step in cortical development." Dev Dyn **236**(3): 696-705.

Chen, H., X. Chen, F. Ye, W. Lu and X. Xie (2009). "Symmetric division and expression of its regulatory gene Numb in human cervical squamous carcinoma cells." Pathobiology **76**(3): 149-154.

Corallini, S., S. Fera, L. Grisanti, I. Falciatori, B. Muciaccia, M. Stefanini and E. Vicini (2006). "Expression of the adaptor protein m-Numb in mouse male germ cells." Reproduction **132**(6): 887-897.

Dho, S. E., M. B. French, S. A. Woods and C. J. McGlade (1999). "Characterization of four mammalian numb protein isoforms. Identification of cytoplasmic and membrane-associated variants of the phosphotyrosine binding domain." J Biol Chem **274**(46): 33097-33104.

Dooley, C. M., J. James, C. Jane McGlade and I. Ahmad (2003). "Involvement of numb in vertebrate retinal development: evidence for multiple roles of numb in neural differentiation and maturation." J Neurobiol **54**(2): 313-325.

Gao, Z., F. L. Chi, Y. B. Huang, J. M. Yang, N. Cong and W. Li (2011). "Expression of Numb and Numb-like in the development of mammalian auditory sensory epithelium." Neuroreport **22**(2): 49-54.

Huang, S.-C., L. V. Vu, F. H. Yu, D. T. Nguyen and E. J. Benz (2021). "Multifunctional protein 4.1R regulates the asymmetric segregation of Numb during terminal erythroid maturation." Journal of Biological Chemistry **297**(3).

Kahles, A., K. V. Lehmann, N. C. Toussaint, M. Huser, S. G. Stark, T. Sachsenberg, O. Stegle, O. Kohlbacher, C. Sander, N. Cancer Genome Atlas Research and G. Ratsch (2018). "Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients." Cancer Cell **34**(2): 211-224 e216.

Kim, K. K., J. Nam, Y. S. Mukouyama and S. Kawamoto (2013). "Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development." J Cell Biol **200**(4): 443-458.

Langer, W., F. Sohler, G. Leder, G. Beckmann, H. Seidel, J. Grone, M. Hummel and A. Sommer (2010). "Exon array analysis using re-defined probe sets results in reliable identification of alternatively spliced genes in non-small cell lung cancer." BMC Genomics **11**: 676.

Misquitta-Ali, C. M., E. Cheng, D. O'Hanlon, N. Liu, C. J. McGlade, M. S. Tsao and B. J. Blencowe (2011). "Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer." *Mol Cell Biol* **31**(1): 138-150.

Moran, T. B., L. B. Goldberg, S. L. Serviss and L. T. Raetzman (2011). "Numb deletion in POMC-expressing cells impairs pituitary intermediate lobe cell adhesion, progenitor cell localization, and neuro-intermediate lobe boundary formation." *Mol Endocrinol* **25**(1): 117-127.

Park, E., Z. Pan, Z. Zhang, L. Lin and Y. Xing (2018). "The Expanding Landscape of Alternative Splicing Variation in Human Populations." *Am J Hum Genet* **102**(1): 11-26.

Petersen, P. H., K. Zou, J. K. Hwang, Y. N. Jan and W. Zhong (2002). "Progenitor cell maintenance requires numb and numblike during mouse neurogenesis." *Nature* **419**(6910): 929-934.

Rasin, M. R., V. R. Gazula, J. J. Breunig, K. Y. Kwan, M. B. Johnson, S. Liu-Chen, H. S. Li, L. Y. Jan, Y. N. Jan, P. Rakic and N. Sestan (2007). "Numb and Numbl are required for maintenance of cadherin-based adhesion and polarity of neural progenitors." *Nat Neurosci* **10**(7): 819-827.

Tapial, J., K. C. H. Ha, T. Sterne-Weiler, A. Gohr, U. Braunschweig, A. Hermoso-Pulido, M. Quesnel-Vallieres, J. Permanyer, R. Sodaei, Y. Marquez, L. Cozzuto, X. Wang, M. Gomez-Velazquez, T. Rayon, M. Manzanares, J. Ponomarenko, B. J. Blencowe and M. Irimia (2017). "An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms." *Genome Res* **27**(10): 1759-1768.

Verdi, J. M., A. Bashirullah, D. E. Goldhawk, C. J. Kubu, M. Jamali, S. O. Meakin and H. D. Lipshitz (1999). "Distinct human NUMB isoforms regulate differentiation vs. proliferation in the neuronal lineage." *Proceedings of the National Academy of Sciences* **96**(18): 10472-10476.

Yoshida, T., A. Tokunaga, K. Nakao and H. Okano (2003). "Distinct expression patterns of splicing isoforms of mNumb in the endocrine lineage of developing pancreas." *Differentiation* **71**(8): 486-495.

Zhang, Y., S. E. Dho, K. Othman, C. D. Simpson, J. Lapierre, A. Bondoc and C. J. McGlade (2022). "Numb exon 9 inclusion regulates Integrinbeta5 surface expression and promotes breast cancer metastasis." *Oncogene* **41**(14): 2079-2094.

Zhong, W., M. M. Jiang, G. Weinmaster, L. Y. Jan and Y. N. Jan (1997). "Differential expression of mammalian Numb, Numblike and Notch1 suggests distinct roles during mouse cortical neurogenesis." *Development* **124**(10): 1887-1897.