Transcriptomics of mussel transmissible cancer MtrBTN2 suggests accumulation of multiple cancerous traits and oncogenic pathways shared among bilaterians

E.A.V. Burioli[1], M. Hammel[1,2], E. Vignal[1], J. Vidal-Dupiol[1], G. Mitta[3], F. Thomas[4], N. Bierne[2], D. Destoumieux-Garzón[1], G.M. Charrière[1]*

[1]IHPE, Univ Montpellier, CNRS, IFREMER, Univ Perpignan Via Domitia, Montpellier, France;

[2]ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France;

[3]IFREMER, UMR 241 Écosystèmes Insulaires Océaniens, Labex Corail, Centre Ifremer du Pacifique, Tahiti, Polynésie française;

[4]CREEC/CANECEV (CREES), MIVEGEC, Unité Mixte de Recherches, IRD 224–CNRS 5290–Université de Montpellier, Montpellier, France;

*, erikaastrid.burioli@gmail.com

Abstract

Transmissible cancer cell lines are rare biological entities giving rise to diseases at the crossroads of cancer and parasitic diseases. These malignant cells have acquired the amazing capacity to spread from host to host. They have been described only in dogs, Tasmanian devils and marine bivalves. The *Mytilus trossulus* Bivalve Transmissible Neoplasia 2 (MtrBTN2) lineage has even acquired the capacity to spread inter-specifically between marine mussels of the *Mytilus edulis* complex worldwide. To identify the oncogenic processes underpinning the biology of these atypical cancers we performed transcriptomics

of MtrBTN2 cells. Differential expression, enrichment, protein-protein interaction network, and targeted analyses were used. Overall, our results suggest the accumulation of multiple cancerous traits that way be linked to the long-term evolution of MtrBTN2. We also highlight that vertebrate and lophotrochozoan cancers could share a large panel of common drivers, which supports the hypothesis of an ancient origin of oncogenic processes in bilaterians.

26 **1. Introduction**

27 Transmissible cancers are rare and fascinating biological entities because they have evolved
28 the ability to overcome the physical and immunological barriers of the host to become
29 contagious and to spread between animals by direct transfer of cancer cells, thus behaving
30 like parasites (1,2). They have been reported in only two vertebrate species so far, in dogs
31 and in Tasmanian devils. The transmissibility of cancer cells was first demonstrated in the
32 Canine Transmissible Veneral Tumor (CTVT) (3,4), which is thought to have originated more
33 than 4,000 years ago (5). CTVT has since spread across continents, probably together with
34 humans, through coitus and oral contact between dogs. Tasmanian Devil Facial Tumor
35 Disease (DFTD) is the other well-known transmissible cancer that affects vertebrates. Two
36 independent cancerous lineages DFT1 and DFT2 have been described to date and have
37 emerged less than fifty years ago (6). Infection occurs via bites, a common behaviour in
38 Tasmanian devil social interactions (7-9). As a result, cancer transmission has devastated
39 much of the devil population, with some local populations declining by as much as 80% in
40 the geographical areas most affected by DFTD (10-13). However, the highest number of
41 transmissible cancer lineages identified to date has been found in bivalve molluscs (8
42 different lineages) and these multiple lineages have been grouped under the term Bivalve
43 Transmissible Neoplasia (BTN) (14-17). BTNs usually transmit between individuals of the
44 same species, but some lineages have crossed the species barrier and are circulating in
45 populations of related host species (15-17). Indeed, one specific lineage of BTN, called
46 *Mytilus trossulus* Bivalve Transmissible Neoplasia lineage 2 (MtrBTN2), has emerged in a *M.*
47 *trossulus* founder host and has since spread to other *Mytilus* species populations (*M.*

48 *trossulus*, *M. edulis*, *M. galloprovincialis*, and *M. chilensis*) across several continents—South
49 America, Asia, and Europe (15,16,18-21).
50 Transmissible cancer lineages must have emerged from a first neoplastic transformation in a
51 founder host and then evolved specific phenotypes to become a new type of contagious
52 etiologic agent. In fact, we have previously shown that conventional non-transmissible
53 cancers also occur in *Mytilus* bivalves (20). MtrBTN2 cells are found circulating in the
54 haemolymph together with haemocytes, the mussel's immune cells. Like haemocytes, they
55 are able to infiltrate connective tissues of various organs (19,22). As the disease progresses,
56 MtrBTN2 cells outgrow the host cells and progressively replace almost all the circulating cells
57 present in the haemolymph and disseminate in the tissues. They have a characteristic
58 morphology of rounded and basophilic cells with a high nucleus-to-cell ratio and they are
59 polyploid (19). Therefore, like other haemolymphatic cancers (20), they are easily diagnosed
60 by histological/cytological observation or flow cytometry of haemolymph (19,23). This initial
61 diagnosis needs to be complemented by genetic analysis to distinguish MtrBTN2 from other
62 transmissible cancers and conventional haemolymphatic cancers (16,24). MtrBTN2 cells
63 proliferate very quickly with a mean doubling time of ~ 3 days (24). In addition, BTNs must
64 be able to survive in the extra-host environment long enough to infect a new host (24,27).
65 We have shown that MtrBTN2 cells can survive at least for 3 days without mortality and up
66 to 8 days in seawater (24). However, the molecular pathways related to the malignancy of
67 BTNs have not been characterised. Based on the phenotypic traits observed in MtrBTN2
68 cells, such as high proliferative activity, genomic abnormalities such as polyploidy,
69 supermetastatic ability, and extended cell survival capabilities, it is likely that multiple
70 molecular pathways contribute to the malignancy and long-term persistence of this lineage.

71 Among the most common and most studied functional capabilities observed in human
72 cancer cells are the sustaining of proliferative signalling, the evasion from growth
73 suppressors, the resistance to cell death, the replicative immortality, the activation of
74 invasion and metastasis, the reprogramming of cellular metabolism, and the avoidance of
75 immune destruction (25). The acquisition of these capabilities is ensured by the activation or
76 inactivation of specific signalling pathways. Among these, Sanchez-Vega et al. (26)
77 highlighted ten oncogenic signalling pathways that are most frequently altered in most
78 human cancers (HIPPO, MYC, NOTCH, NRF2, PI3K, RTK/RAS, TGFβ, P53, WNT, and cell cycle),
79 and are mainly involved in promoting cell proliferation. These conclusions were drawn from
80 extensive data on hundreds of human cancers obtained by transcriptomic, genomic and
81 epigenomic analyses.

82 Measuring relative levels of gene expression has been a key approach to identify genes and
83 biological pathways associated with the cancerous process and functional adaptations of
84 cancers (28). In recent years, RNA sequencing (RNA-seq) has emerged as a fast, cost-
85 effective, and robust technology to address various biological questions (29,30).
86 Transcriptome analyses allow to link cellular phenotypes to their molecular underpinnings.
87 In the context of cancers, these links provide an opportunity to dissect the complexity of the
88 cancer biological adaptations. For non-model organisms and in the absence of a suitable
89 reference genome, as is the case for *M. trossulus*, RNA-seq is used to reconstruct and
90 quantify *de novo* transcriptomes (31). Differential gene expression analysis is then carried
91 out to compare the effect of treatments or conditions on gene expression. To date, although
92 several transmissible neoplasia have been described in marine bivalves, transcriptome-wide
93 functional studies are still lacking. However, Bruzos et al. (pre-print 2022, 32) have

94 performed a comparative quantitative transcriptomic study between cancer cells and
95 various healthy tissues in cockle BTNs, which supports their status as haemocyte-derived
96 marine leukaemias.

97 Here, we performed a deep sequencing of cancer cell transcriptomes to investigate the gene
98 expression profile of MtrBTN2. We sequenced mRNA from circulating cells in both healthy
99 and MtrBTN2-affected *M. edulis*, and since that the founder host species was *M. trossulus*,
100 we included haemocytes from this species as a control. Analysis of differentially expressed
101 genes unveiled some functional characteristics of this unusual cancer and highlighted
102 potential cancer drivers common to all bilaterians.

103

104 **2. Results**

105 MtrBTN2 cells were collected from the haemolymph of *M. edulis* mussels reared in the
106 English Channel (Normandy, France). Only individuals with more than 95% of circulating
107 neoplastic cells in their haemolymph were used for MtrBTN2 cell collection. *M. edulis*
108 haemocytes were collected from individuals from the same geographical location but tested
109 negative for MtrBTN2 by cytology and qPCRs targeting a sequence in the EF1 gene specific
110 for the *M. trossulus* species and a MtrBTN2 specific sequence in the mitochondrial COI gene
111 (16,24). *M. trossulus* haemocytes were collected from the haemolymph of wild *M. trossulus*
112 found in the Barents Sea (Kola Bay, Russia), as this species is absent from the French coasts.
113 All *M. trossulus* mussels were positive for the qPCR targeting the *M. trossulus* specific
114 sequence, but negative by cytology and qPCR for MtrBTN2. We did not observe any other
115 pathogen by histology in the nine individuals used for RNA-Seq analysis. Transcripts from

haemolymph samples were sequenced for each condition: MtrBTN2-positive (i.e. cancerous) *M. edulis* (CAN 1-3), MtrBTN2-negative *M. edulis* (EDU 1-3) and MtrBTN2-negative *M. trossulus* (TRO 1-3). All RNA-seq reads were used to assemble *de novo* a pantranscriptome from the 9 samples (CAN 1-3, EDU 1-3, and TRO 1-3). The graphical representation of the experimental setup and bioinformatic analyses is shown in Figure 1.

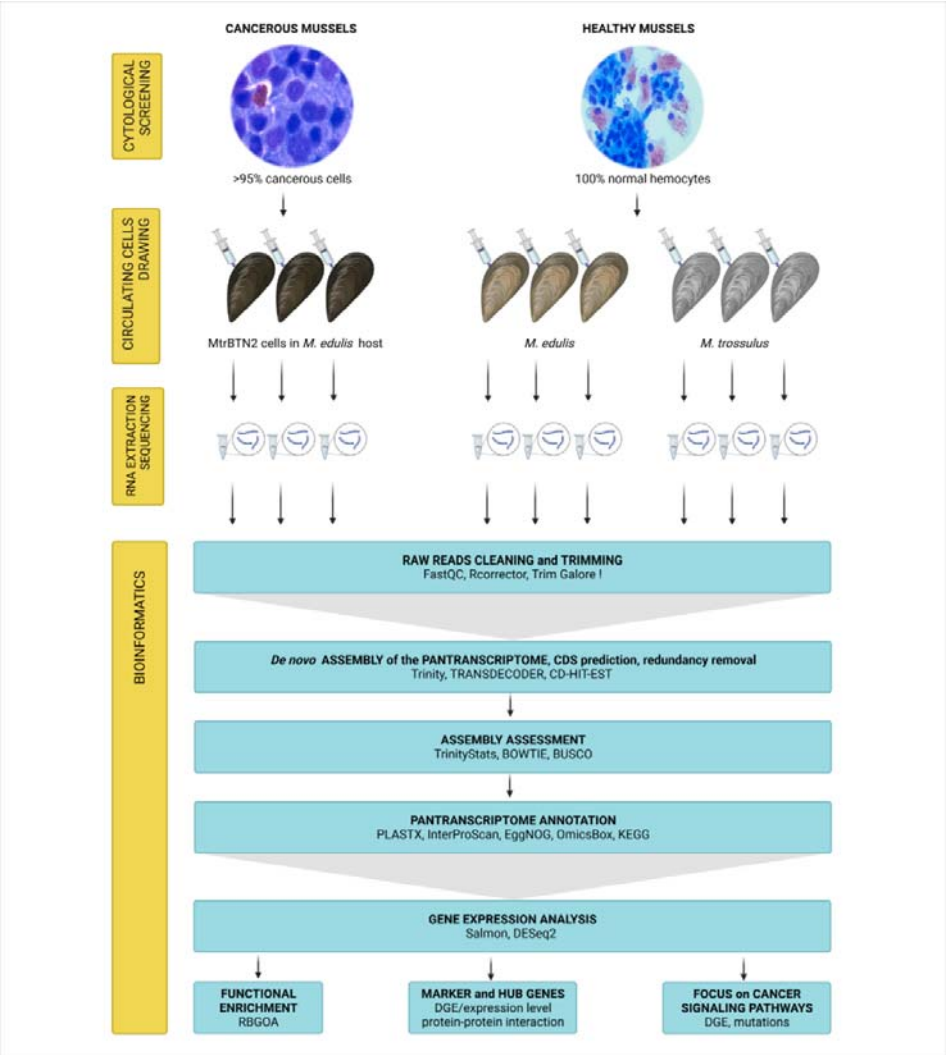


Figure 1: Graphical representation of the experimental design and RNA-Seq data analysis. DGE: differential gene expression analysis. Microphotographs of cancerous and healthy circulating cells are on the same scale.

124 **2.1 Pantranscriptome assembly and annotation**

125 The raw assembly contained 2,149,788 transcripts. After filtering, we retained a total of
126 138,071 transcripts consisting of predicted coding sequences (CDS). The pantranscriptome
127 completeness assessment revealed that 95.6% of the highly conserved single-copy metazoan
128 genes (n = 954) were present in full length as a single (81.2%) or duplicated (14.4%) copies.
129 The mean remapping rate was 53.38 ± 0.58% with negligible differences between samples
130 (Kruskal-Wallis, H(2)=1.69, p=0.43), which is acceptable for a remapping on CDS (33). Only
131 5826 transcripts showed no blast similarity or a conserved domain. The OmicsBox program
132 v2.0 (34,35) assigned GO terms to 131,346 transcripts, but ultimately 41.39% (57,154) of
133 transcript annotations were retained after computing a filtration step based on a minimum
134 annotation score (Supplementary Table S1). The assembly and annotation metrics are shown
135 in Table 1. Raw read sequences coming from RNA-seq were deposited in the NCBI BioProject
136 database (https://www.ncbi.nlm.nih.gov/bioproject/) with links to BioProject accession
137 number PRJNA749800. Individual SRA numbers are listed in Supplementary Table S2.

Attributes	Cancerous individuals			Healthy individuals					
	MtrBTN2			*Mytilus edulis*			*Mytilus trossulus*		
	CAN1	CAN2	CAN3	EDU1	EDU2	EDU3	TRO1	TRO2	TRO3
Clean read pairs x10^{6}	83,69	82,83	82,05	83,88	81,93	98,33	81,57	76,67	84,82
GC%	36	36	36	37	37	37	36	36	36
Transcriptome statistics									
	Raw number of transcripts						2,149,788		
	Final number of transcripts						138,071		
	GC%						38		
	Total assembled bases						97,665,057		
	Contigs N50 (bp)						1,038		
	Median contig length (bp)						439		
	Average contig length (bp)						745		
	BUSCO % (*Metazoa, n=954*)						95,6		
		Complete single-copy					167		
		Duplicated					745		
		Fragmented					34		
		Missing					8		
Back alignment on pantranscriptome %	53,53	53,92	53,48	53,88	53,83	52,83	53,70	52,18	53,14
Annotation statitics									
	Transcripts with GO associated						131,346		
	Transcrips annotated*						57,154		
	No-blast, no Interproscan hits found						5,826		

Table 1: Assembly and annotation metrics. The pantranscriptome was obtained from the mRNA sequencing of circulating cells from three *Mytilus edulis*, three *M. trossulus,* and three MtrBTN2-affected *M. edulis* at a late stage of the disease (>95% of cancer cells).*after applying the OmicsBox annotation rule (e-value hit filter of 1.10^{-3}, annotation cutoff of 55, a HSP-Hit coverage of 60%).

2.2 The specific transcriptomic profile of MtrBTN2 cells

Principal component analysis (PCA) based on the normalised count matrix (RLE values) discriminated the three groups of samples (CAN, EDU, TRO) well (Figure 2). These first results indicate that the transcriptomic profile of MtrBTN2 differs from the transcriptomic profile of healthy *Mytilus* haemocytes, irrespective of their species or geographical origin.

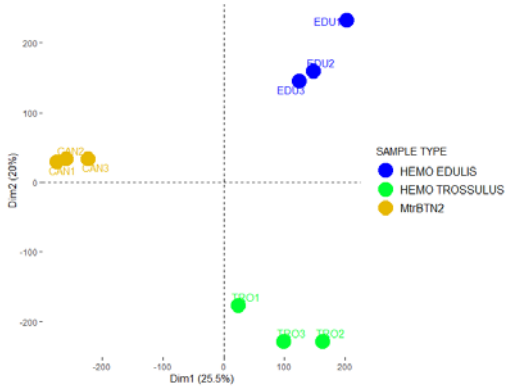
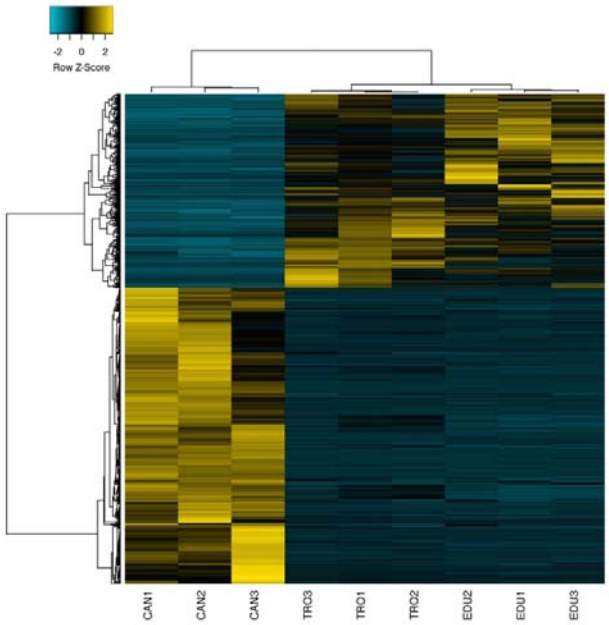


Figure 2: Principal component analysis (PCA) based on the normalised count matrix (RLE values). The first axis discriminates MtrBTN2 cells from hemocytes.

We investigated differentially expressed transcripts in MtrBTN2 cells (CAN 1-3) by comparing their abundance with *M. edulis* (EDU 1-3) and *M. trossulus* haemocytes (TRO 1-3) using DESeq2 v1.34 (36). These comparative results were then validated by RT-qPCRs (see sections 5.3 and 5.5) targeting 17 key-gene transcripts and confirmed in 10 additional haemolymph samples (CAN 7-11, EDU 7-11) (see Supplementary Table S3; Supplementary Figure S1 and S2). A total of 4,356/138,071 transcripts were significantly differentially expressed; 2,632 were more expressed and 1,724 were less expressed in MtrBTN2 cells than in haemocytes (see Log2 fold change (Log2FC) and adjusted p-values (FDR) in Supplementary Table S4). We further validated the significance of differential expression for 4,254 transcripts by running QuantEval (37), which builds connected components from the assembled transcripts based on sequence similarity and evaluates the quantification results for each connected component (Supplementary Tables S4 and S5, Supplementary Figure S3). Indeed, for 102 transcripts, the significance of the differential expression was not confirmed at the component level. A hierarchical clustering of DETs is shown on a log2 centred-heatmap (Figure 3). Although the results were less contrasted with RNA-seq data obtained from the

166 whole flesh of other 6 additional individuals coming from a separate study (CAN 4-6, EDU 4-
167 6) (see sections 5.3 and 5.4; Supplementary Table S3) than from haemolymph cells, the
168 principal component analysis of the transcript abundances also confirmed a differential
169 clustering between MtrBTN2-affected and MtrBTN2-free mussels (Supplementary Figure S4).



170

171 **Figure 3:** Heatmap (log2 centred) with hierarchical clustering of the 4254 differentially expressed transcripts at
172 a FDR <0.05.

173 Among the most differentially expressed transcripts (DET) (with log2FC > 5), we identified six
174 top-expressed (Relative Log Expression (RLE) > 500) transcripts of the genes *CYP11A1*,
175 *THBS1*, *SELP*, *ALOX5*, *BMP2*, and *ADAMTS1* (Figure 4). These transcripts were also confirmed
176 among the most differentially and top-expressed in the whole flesh of the additional 3
177 individuals (CAN 4-6) (see sections 5.3 and 5.5; Supplementary Table S3; Supplementary
178 Figure S5). Indeed, these genes can be considered as specific for the malignant state and
179 have been described to be involved in human tumorigenesis and metastasis (Table 2).

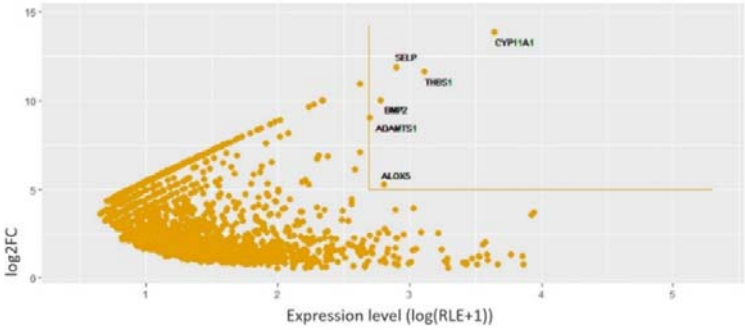


Figure 4: Scatter plot showing genes specific of the malignant state with y=log2FC, x=log(RLE+1). Lines delimit the areas above log2FC values > 5 and RLE > 500. Six transcripts of genes (*CYP11A1*, *THBS1*, *SELP*, *ALOX5*, *BMP2*, and *ADAMTS1)* exceeded these values and can be considered as specific of the malignant state.

gene	Product	Function	Biological processes	References
CYP11A1	Cholesterol side-chain cleavage enzyme	- catalysis of cholesterol to pregnenolone convertion	- modulation of immune response (glucocorticoids) - sexual development and gametogenesis (androgens and estrogens)	Hu et al., 2010 [38]
THBS1	Thrombospondin-1	- cell-to-cell and cell-to-matrix interactions - major activator of TGF-β to its mature form	- chemotactic gradient for immune cells participating in inflammatory response during acute phase - induction macrophage polarization to the M2 anti-inflammatory phenotype	Murphy-Ullrich and Poczatek, 2000 [39] Letterio and Roberts, 1998 [40] Ashcroft, 1999 [41]
SELP	P-selectin	- vascular adhesion molecules	- extravasation of circulating metastatic cells - inflammation - induction macrophage polarization to the M2 anti-inflammatory phenotype	Läubli and Borsig, 2010 [42] Fabricius et al., 2021 [43]
ALOX5	5-Lipoxygenase	- synthesis of leukotrienes	- inflammation - proliferation of cancerous cells - apoptosis inhibition of cancerous cells	Anderson et al., 1998 [44] Bishayee et al., 2013 [45]
BMP2	Bone morphogenetic protein 2	- growth factor - cell differentiation	- stemness maintenance - induction macrophage polarization to the M2 anti-inflammatory phenotype	Lee et al., 2013 [46]
ADAMTS1	A disintegrin and metalloproteinase with thrombospondin motifs 1	- matrix proteolytic enzyme	- ECM dismantling - induction macrophage polarization to the M2 anti-inflammatory phenotype	Redondo-García et al., 2021 [47] Rucci et al., 2011 [48]

Table 2: Top-expressed transcripts of genes in MtrBTN2 cells (log2FC > 5 and RLE > 500). Most are involved in inflammation and immune response.

They are associated with inflammation, immune processes, and extracellular matrix (ECM) disruption (38-48). The exact role of some of these genes (*THBS1, SELP, ALOX5,* and *BMP2*) during tumorigenesis is still under debate, mainly for their involvement in the inflammatory process (49,50). However, *THBS1* expression is increased in many cancers promoting invasion and metastasis (51-54). *SELP* and *ALOX5* are highly expressed in some tumours and

192 cancer cell lines, inducing cancer cell proliferation (55-58) and inhibiting apoptosis (44,45).
193 There are also conflicting data regarding the effect of *BMP2* on cancer (59). Nevertheless,
194 most studies suggest that *BMP2* promotes metastatic progression and tumorigenesis.
195 Although mollusc immunity against cancer has never been studied so far, most of these
196 genes could be involved in modulating the host immune response. Some of them, such as
197 *THBS1*, *SELP*, *BMP2*, and *ADAMTS1,* are involved in the induction of macrophage polarisation
198 towards the anti-inflammatory M2 phenotype in mammals and their role in the interactions
199 of MtrBTN2 with the host immune response deserves further attention to better understand
200 the host invasion process.

201 **2.3 Dysregulated biological processes in MtrBTN2**

202 To identify the key biological processes altered in the MtrBTN2 phenotype, we performed a
203 GO_MWU rank-based enrichment analysis on the DETs (for protein descriptions and gene
204 IDs, see Supplementary Table S4). A biological process is considered to be altered if DETs are
205 significantly overrepresented (enriched) within the total set of transcripts related to that
206 process. The complete list of transcripts clustered by Gene Ontology (GO) term is provided in
207 the Supplementary Table S6. The GO_MWU analysis revealed several biological functions
208 that were differentially regulated in MtrBTN2 cells compared to normal haemocytes (Figure
209 5).

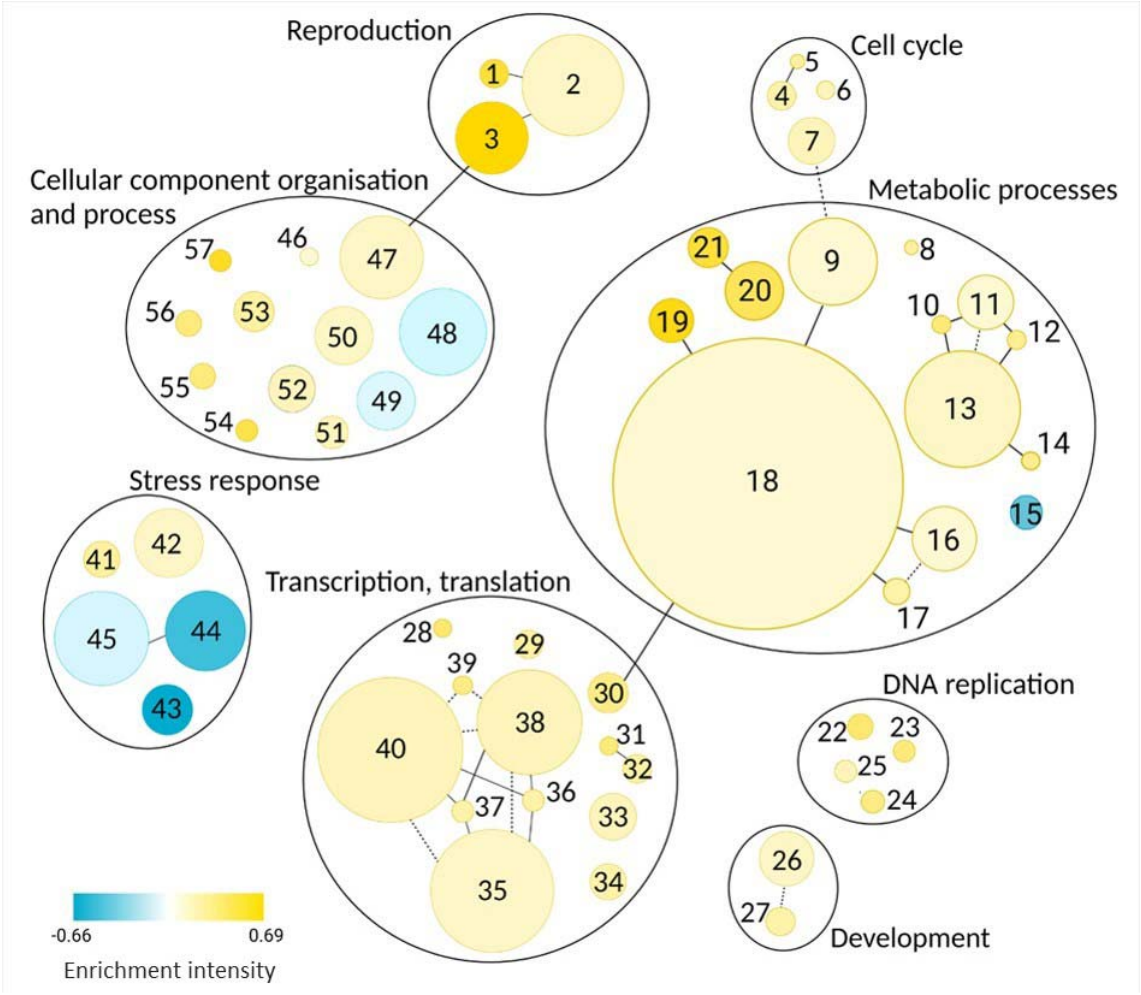


Figure 5: Biological processes dysregulated in transmissible neoplasia (p-value < 0.05): **1.**Meiosis I; **2.**Cellular process involved in reproduction; **3.**Sperm motility; **4.**Cytokenetic process; **5.**Contractil ring contraction; **6.**Mitotic spindle elongation; **7.**Regulation of cell cycle process; **8.**Protein deacylation; **9.**Proteolysis; **10.**Adenosine metabolic process; **11.**Nucleoside phosphate biosynthetic process; **12.**IMP metabolic process; **13.**Carbohydrate derivative metabolic process; **14.**Dolichol-linked oligosaccharide biosynthetic process; **15.**Tricarboxylic acid cycle; **16.**Peptidyl-amino acid modification; **17.**Histone modification; **18.**Protein metabolic process; **19.**Protein deubiquitination; **20.**Nucleobase metabolic process; **21.**Pyrimidine nucleobase metabolic process; **22.**DNA replication initiation; **23.**DNA strand elongation; **24.**DNA replication, synthesis of RNA primer; **25.**DNA-dependent DNA replication; **26.**Regulation of growth; **27.**Cell fate determination; **28.**Cis assembly of pre-catalytic spliceosome; **29**.RNA phosphodiester bond hydrolysis; **30.**Mitochondrial translation;

221 **31.**Transcription initiation from RNA polymerase I promoter; **32.**DNA-templated transcription, initiation;
222 **33.**Translational elongation; **34.**Spliceosomal snRNP assembly; **35.**mRNA processing; **36.**tRNA splicing, via
223 endonucleolytic cleavage and ligation; **37.**tRNA wobble base modification; **38.**tRNA metabolic process;
224 **39.**Deadenylation-dependent decapping of nuclear-transcribed mRNA; **40.**ncRNA metabolic process;
225 **41.**Response to oxidative stress; **42.**Double-strand break repair; **43.**Cellular response to amino acid starvation;
226 **44.**Response to bacterium; **45.**Defense response; **46.**Actin filament based process; **47.**Microtubule-based
227 movement; **48.**Cell surface receptor signalling pathway; **49.**Anion transport; **50.**Golgi vesicle transport;
228 **51.**Ribosomal small subunit assembly; **52.**Ribosomal large subunit assembly; **53.**rRNA-contain RNP complex;
229 **54.**Protein secretion; **55.**Protein targeting to ER; **56.**Mitochondrial respiratory chain complex assembly;
230 **57.**Protein insertion into ER membrane. The color intensity is proportional to the enrichment intensity,
231 expressed as the ratio between the number of transcripts that were significantly over- (yellow) or down-
232 expressed (blue) compared with the total number of transcripts in the process. The disk area is proportional to
233 the number of DETs linked to the process (from 2 to 191 transcripts). Lines connecting two functions represent
234 shared transcripts (solid line means that all transcripts in the smaller family are part of the largest family;
235 dotted line means that only some transcripts are shared).

236 ***Proliferation-related functions are altered***

237 Some functions consistent with a high proliferative activity are altered in MtrBTN2 cells. The
238 "cell cycle" family (4.Cytokenetic process; 5.Contractil ring contraction; 6.Mitotic spindle
239 elongation; 7.Regulation of cell cycle process, in Figure 5) is enriched in MtrBTN2 cells, as
240 well as "DNA replication" (22.DNA replication initiation; 23.DNA strand elongation; 24.DNA
241 replication, synthesis of RNA primer; 25.DNA-dependent DNA replication, in Figure 5). More
242 specifically, negative regulation of the cell cycle progression appeared to be reduced: *ASAH2,*
243 which inhibits the ceramide mediation of the cell cycle arrest (60), was more highly
244 expressed, whereas *PPP2R5B,* which inhibits cell cycle progression (61), was less expressed
245 in MtrBTN2. Moreover, transcripts of genes promoting the cell cycle progression were more

246 abundant in MtrBTN2 cells, in particular five genes that form the Anaphase Promoting
247 Complex (*ANAPC1, ANAPC4, ANAPC5, CDC26, and CDC23*) and *CCND1* encoding cyclin D1,
248 which promotes progression through the G1-S phase of the cell cycle (62). *TERT,* which is
249 essential for the chromosome telomere replication in most eukaryotes and is particularly
250 active in progenitor and cancer cells (63) was also more expressed in MtrBTN2 cells. By
251 maintaining telomere length, telomerase removes a major barrier to cellular longevity,
252 allowing unlimited proliferation driven by oncogenes (64). In addition, in the "Development"
253 family, we found a higher abundance of transcripts of genes involved in promoting cell
254 growth, such as *TTK* (65), *METAP2* (66), *THBS1* (activator of TGF-β) (39), *MEAK7* (67) and
255 *LAMTOR1* (68). Transcripts of several genes that act as growth inhibitors were less abundant
256 in MtrBTN2, as often seen in cancer contexts, such as *FAM107A* (69) and *LTBP3* (70), which
257 keeps TGF-β in a latent state.

258 ***Metabolic pathways are modified***

259 Many metabolic processes were shown to be altered (Figure 5 and Supplementary Table S6).
260 Nucleotide metabolism in MtrBTN2 cells was characterised by a higher expression of more
261 than 25 transcripts of genes involved in the *de novo* and salvage biosynthetic pathways of
262 purines and pyrimidines and in the homeostasis of cellular nucleotides. The metabolism of
263 protein and carbohydrate derivatives was also modified. The enrichment concerned
264 transcripts of genes encoding ribosomal proteins, protein turnover and localisation
265 (especially deubiquitination), proteolysis, peptidyl amino-acid modification, histone
266 acetylation, and lipid and protein glycosylation. In addition to *MEAK7* and *LAMTOR1*,
267 transcripts of eight genes (*BMP2*, *KICS2*, *ITFG2*, *MIOS*, *NPRL2*, *RagA, RagD, and SEC13*)
268 involved in the mTORC1 amino-acid sensing pathway (43.Cellular response to amino acid

269 starvation, in Figure 5) (71) were differentially expressed. Interestingly, in addition to the
270 role of mTORC1 in the regulation of autophagy, mTORC1 plays a pivotal role as a master
271 regulator of protein, lipid, and nucleic acid metabolism changes that have been reported in
272 many different cancer cells (72-75). Several transcripts of genes involved in organelle
273 activities and related to protein metabolism (anabolism, catabolism, post-translational
274 modifications, folding, and secretion) were also more highly expressed in MtrBTN2.

275 The tricarboxylic acid cycle (TCA) also appeared to be differentially regulated in MtrBTN2,
276 with a lower expression of *DLST* (a component of the 2-oxoglutarate dehydrogenase
277 complex), *SDHD* and *NNT*, and a higher expression of *SUCLA2*. The TCA cycle is critical for the
278 generation of cellular energy and precursors for biosynthetic pathways (76). By screening
279 the expression of 16 glycolysis-related genes, we found that *PFK*, *PKM*, *SLC2A1, and SLC2A3*
280 were more expressed; however, *PGK1*, *TPI1*, and *ALDOA* were less expressed than in control
281 haemocytes. These conflicting results did not allow us to conclude on how the regulation of
282 glycolytic activities is altered in transmissible neoplasia cells. However, these results suggest
283 major changes in nucleotide, amino-acid and energy metabolism in MtrBTN cells as often
284 seen in other cancers (77).

285 ***DNA repair systems are dysregulated***

286 MtrBTN2 cancer was found to express at higher levels a significant panel of transcripts of
287 genes involved in the DNA Homologous Recombination (HR) (42.Double-strand break repair,
288 in Figure 5), such as *BABAM1*, *ZSWIM7*, *RAD51L*, *RAD54L*, *MRE11*, *MCM9*, and *TONSL*. The
289 DNA helicase MCM9 is also involved in the repair of inter-strand crosslinks (78), as is MUS81
290 (79), which is also more expressed in MtrBTN2 cells. This type of damage is a double-edged

291 sword, as double-strand breaks can induce cell death if not efficiently repaired, but
292 inefficient or inappropriate repair can lead to mutation, gene translocation and cancer (80).
293 Due to their high proliferation rate and increased metabolic activities, cancer cells are under
294 a huge replication stress (81), which may emphasise the need for these long-lived cancer
295 lineages to activate DNA repair to persist on time. Cells are able to repair double-strand
296 breaks through two major pathways: non-homologous end joining and HR (82). HR is critical
297 for re-establishing replication forks at the sites of damage during S and G2 phases of the cell
298 cycle. However, when combined with the depletion of cell cycle checkpoints and apoptosis,
299 this mechanism can lead to genomic instability. As the ploidy of MtrBTN2 appears to be
300 highly abnormal (varying from 8N to 18N, (19)), further investigations will be required to
301 determine whether the dysregulation of these DNA repair pathways favours recombination
302 between homologous chromosomes and the accumulation of chromosomal abnormalities,
303 or whether these pathways are activated in order to limit DNA damage. Interestingly,
304 upregulation of double-stranded DNA break repair via the HR pathway has also been found
305 in DFTD, in addition to repair *via* non-homologous end joining (83).

306 ***Interaction with ECM is affected***

307 The expression levels of transcripts of seven genes encoding integrin subunits (48.Cell
308 surface receptor signalling pathway, in Figure 5), a major class of transmembrane
309 glycoproteins that mediate cell-matrix and cell-cell adhesion, were altered (Supplementary
310 Table S6). As integrins are mostly less expressed in MtrBTN2 cells, this suggests that the
311 ability of these cells to interact with ECM components is profoundly altered, which is often
312 involved at multiple stages of the cancer processes (84).

313 ***Cell fate determination pathways are modified and soma-to-germline transition may have***
314 ***occurred***

315 Major cell fate pathways appear to be modified (Figure 5 and Supplementary Table S6). The
316 expression of transcripts of six genes involved in the Notch and Wnt pathways was altered.
317 These changes may drive dedifferentiation processes often observed in aggressive cancer
318 cells or cancer stem cells (85). We report the results of the focused analysis performed on
319 these two signalling pathways in section 2.5. In addition, three biological processes clustered
320 under GO terms related to "reproduction" were enriched (1.Meiosis I; 2.Cellular process
321 involved in reproduction; 3.Sperm motility, in Figure 5). Nineteen transcripts more expressed
322 in MtrBTN2 cells gathered under the "sperm motility" process and encode dynein assembly
323 factors, dynein chains, and cilia and flagella associated proteins. Moreover, transcripts of
324 genes involved in meiosis such as *MSH5* (86), *SYCP3* (87) and *TEX11* (88), were more
325 abundant in MtrBTN2. These unexpected results may reflect a soma-to-germline transition
326 during the oncogenic process, which has been reported in a variety of cancers, and support
327 the hypothesis of profound changes in the differentiation state of MtrBTN2 cells (89,90).
328 Interestingly, when abnormally produced in mitotically dividing cells, the *SYCP3* can impair
329 recombination and drive ploidy changes that affect chromosome segregation in cancer cells,
330 which may be one of the mechanisms that led to the hyperploidy of MtrBTN2 cells (90).

331 ***Immune response is altered***

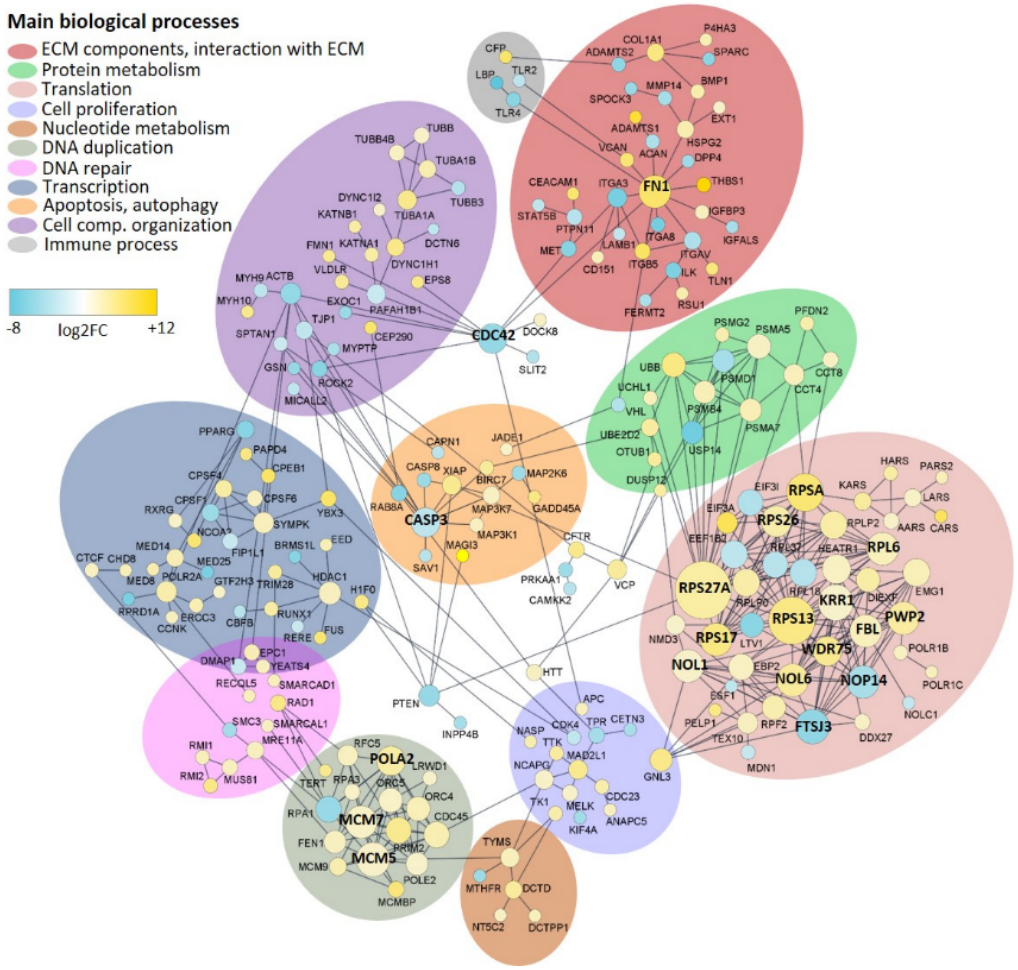
332 One of the most striking differences between MtrBTN2 cells and haemocytes is that
333 biological processes belonging to innate immunity (44.Response to bacterium; 45.Defense
334 response; in Figure 5) are significantly underrepresented in MtrBTN2 cells. Transcripts of
335 genes encoding important mussel antimicrobial peptides and bactericidal proteins such as

336 *Mytilin B*, *Myticin A*, *Myticin B*, *MGD1*, and *BPI* (91-93) were significantly less abundant in
337 MtrBTN2 cells. A similar pattern was also observed for genes encoding several pattern
338 recognition receptors, such as *TLR2* and *TLR4,* which belong to the Toll-Like-Receptor family
339 and are essential for innate immunity against pathogens (94), and the lectin-related pattern
340 recognition molecules *FCN2*, *FIBCD1*, *CLEC4E*, and *CLEC4M*, which are involved in the
341 recognition of pathogens and apoptotic or necrotic cells (95-97). This underrepresentation of
342 biological functions related to host defences suggests that MtrBTN2 cells are
343 immunologically incompetent. During MtrBTN2 disease progression, we also observe a
344 drastic decrease in the number of host haemocytes, which are progressively replaced by
345 circulating MtrBTN2 cells, which can represent over 95% of the circulating cells in cancerous
346 mussels (19,22). It remains to be determined whether MtrBTN2 cells outcompete normal
347 haemocytes in the haemolymph or whether they interfere with haematopoiesis, as seen in
348 some human leukaemias (98). Still, the combination of fewer circulating haemocytes and the
349 immune incompetence of MtrBTN2 cells could have dramatic consequences for the host
350 health, leading to lethal opportunistic systemic infections (99).

351 **2.4 *CASP3*, *FN1*, and *CDC42* are hub genes in the interaction network of MtrBTN2**

352 Protein-protein interaction (PPI) network of the DETs was constructed to evaluate their
353 connectivity and to identify hub genes that may play a critical role in the MtrBTN2
354 phenotype. The complete PPI network is shown in Figure 6. The top 20 genes in the
355 connectivity ranking were found in functions already identified as enriched with GO_MWU,
356 such as translation (most involved in ribosome biogenesis) and DNA replication. Outputs
357 from NetworkAnalyzer v4.4.8 (100) are reported in Supplementary Table S7. Interestingly,
358 this analysis highlighted three genes as major hubs in the PPI network that did not stand out

359 in the GO_MWU analysis, namely *FN1, CDC42,* and *CASP3*. Fibronectin 1, encoded by *FN1*, is
360 a major component of the ECM and plays an important role in cell adhesion, migration,
361 growth and differentiation (101). The abundance of *FN1* transcripts was significantly higher
362 in MtrBTN2 cells (log2FC > 5). CDC42*,* a member of the Rho GTPase family, plays an
363 important role in cell-cell and cell-matrix adhesion, actin cytoskeleon architecture, cell
364 migration, and cell cycle regulation (102); corresponding transcripts were significantly less
365 abundant in MtrBTN2 cells. As Fibronectin 1 and CDC42 are involved in complementary
366 functions and signalling pathways (such as cell-ECM interactions and cell migration), both
367 hub genes were found to be interconnected in the PPI network. In addition, a large number
368 of integrin subunits were found to be dysregulated. Taken together, these results strongly
369 suggest that cell-ECM interactions and cell migration functions are profoundly modified in
370 MtrBTN2 cells, consistent with their observed rounded morphology, poor adhesion
371 properties and f-actin modifications (19,21). The master effector of apoptosis, Caspase 3,
372 encoded by *CASP3*, is also less expressed in MtrBTN2 cells. It plays a central role in the
373 execution phase of cell apoptosis (103), which is often inhibited in cancer cells, allowing
374 them to escape cell death, despite the accumulation of DNA damage and cell cycle
375 dysfunction.



376

377 **Figure 6:** Protein-protein interactions (PPI) network. The size of the disc is proportional to the number of
378 connections. The top 20 Hub proteins are shown in bold. Up-regulated genes are shown in yellow; Down-
379 regulated genes are shown in blue. The color intensity is proportional to the Log2FC value.

380 **2.5 Most key oncogenic signalling pathways are altered in MtrBTN2**

381 Based on the Cancer Genome Atlas project, Sanchez-Vega et al. (26) highlighted that 10
382 signalling pathways are very frequently altered in most human cancers. We found that six of
383 these 10 oncogenic signalling pathways may be altered in MtrBTN2 cancer. These pathways
384 are involved in cell proliferation and differentiation.

385 DETs were found for key genes related to the “Hippo”, “Notch”, “Wnt”, “Myc”, “PI3K”, and
386 “Cell Cycle” signalling pathways (Table 3).

Pathway	Main functions	Gene	Action on pathway	MtrBTN2/hemocytes expression	Log2FC
Hippo	anti-proliferative anti-stem cell self renewal pro-apoptotic	SAV1	activation	Lower	-3,65
Notch	cell proliferation cell fate	JAG1 (2 genes) FBXW7	activation repression	Higher Higher	3.75 and 4.32 1,24
Wnt	cell proliferation stemness maintenance	SFRP1 SFRP3 APC CTNNB1	repression repression repression /	Lower Higher Higher Higher	-3,54 7,38 2,63 1,09
Myc	cell proliferation apoptosis	MAD3	repression	Lower	-2,19
PI3K	cell proliferation cell survival	PTEN PIK3R1	repression repression	Lower Lower	-4,51 -1,68
Cell Cycle		CCND1	activation	Higher	1,89

388 **Table 3: Transcripts of genes differentially expressed in MtrBTN2 among the most frequently altered genes in**
389 **the human oncogenic signalling pathways**. We report their action on the corresponding pathway, their
390 expression status, and the Log2(FC). Two genes encoding JAG1 (confirmed as distinct by QuantEval analysis)
391 were more expressed in MtrBTN2 cells.

392 The Hippo signalling pathway plays a role in inhibiting cell proliferation, controlling cell fate,
393 and promoting apoptosis through the phosphorylation of YAP/TAZ transcription
394 coactivators, their cytoplasmic retention and degradation. *SAV1* is a key activator of the
395 Hippo pathway involved in LATS1/2 phosphorylation, a necessary step for subsequent
396 YAP/TAZ phosphorylation (104). As *SAV1* is less expressed in MtrBTN2 cells, the Hippo
397 pathway may be downregulated or inactivated.
398 The PI3K pathway also showed dysregulation. It is involved in cell metabolism, growth,
399 proliferation, cell-ECM interactions and survival. We observed a lower expression of two
400 genes, *PTEN* and *PIK3R1*, which act as PIK3 antagonists (105). When the pathway is
401 activated, AKT1 is phosphorylated and inhibits the activity of TSC1/2. AKT-mediated

402 phosphorylation of TSC1/2 removes its inhibition of RHEB activity, leading to activation of
403 the mTORC1 complex. Our enrichment analysis (section 2.3) has also highlighted the
404 activation of mTORC1 by the alternative amino-acid sensing pathway. In addition, AKT1 plays
405 an indirect role in the Wnt and Myc pathways by negatively regulating GSK3β through its
406 phosphorylation (106). Thus, the inhibitory effect of unphosphorylated GSK3β on MYC and
407 its contribution in the destruction complex of β-cathenin in the Wnt pathway may be
408 repressed in MtrBTN2.
409 The gene encoding MAD3, an antagonist of MYC for MAX binding (107), was less expressed
410 in MtrBTN2 cells, suggesting that the MYC-MAX complex is favoured and the pathway is
411 activated.
412 Among SFRPs, which act as Wnt pathway inhibitors by binding extracellular WNT ligands,
413 *SFRP1* was less expressed and *SRFP3* more. Moreover, *APC,* which encodes a constitutive
414 protein of the Destruction Complex, was more expressed. APC is a multifunctional protein
415 and is involved, for instance, in the normal compaction of mitotic chromatin (108). We
416 looked for differentially expressed genes among the downstream genes of the Wnt pathway,
417 which transcription is regulated by unphosphorylated CTNNB1 mediation. We found that
418 *CCND1*, *JAG1*, and *FN1* were more expressed suggesting that the Wnt pathway was activated
419 (109). CCND1 promotes cell proliferation and JAG1 is a transmembrane ligand that activates
420 the Notch pathway. However, FBXW7, which inhibits the NOTCH cleavage and its
421 subsequent activation, was more expressed in MtrBTN2 cells. In addition, the *HES1*
422 expression, which is promoted by activated NOTCH, was lower in MtrBTN2 cells. HES1
423 negatively regulates the expression of downstream target genes such as tissue-specific
424 transcription factors (110).

To identify potentially relevant mutations related to these pathways, we screened for the presence of MtrBTN2-specific SNVs in the transcript sequences of key genes. Although the recovery of complete transcript haplotypes was hampered by the polyploidy of cancer cells and the short-read sequencing data, we were able to identify MtrBTN2-specific SNVs, variants that were only observed in the three MtrBTN2 samples (CAN 1, 2, 3) but not in the 6 MtrBTN2-free samples (TRO1-3, EDU1-3). MtrBTN2-specific SNVs were identified in 8 key genes (Table 4; for alignments see Supplementary Files S1 and S2) and were confirmed to be present in the additional cancerous individuals (CAN 4-6).

			Substitution type			
Gene	**Transcript lengtl**	**Pathway**	**Synonymous**	**Missense**	**Total number**	**Missense substitution in a domain**
PIK3CA	3144 bp	PI3K	33	0	33	-
PIK3CB	3159 bp	PI3K	3	0	3	-
PTEN	1395 bp	PI3K	2	0	2	-
AKT	1458 bp	PI3K	1	1	2	1
p53	1797 bp	p53	4	1	5	0
MDM2	1647 bp	p53	21	10	31	0
MYC	1269 bp	Myc	0	2	2	2
FBXW7	2100 bp	Notch	1	-	1	-

Table 4: Genes with transcript sequences carrying MtrBTN2-specific SNVs among the most frequently altered genes in the oncogenic signalling pathways. AKT and MYC carried missense substitutions in a protein domain.

The highest number of MtrBTN2-specific SNVs were found in *PIK3CA* (PI3K pathway), with 33 MtrBTN2-specific SNVs/3144 nucleotides. All these specific SNVs were synonymous substitutions but missense substitutions were present in four genes: *AKT*, *p53*, *MDM2*, and *MYC*. In two of these genes, the missense substitutions were located within protein domains. The missense substitution in the *AKT* sequence was located in the kinase domain, and the two missense substitutions in the *MYC* sequence in the transcriptional activation domain. Interestingly, in the *p53* gene which encodes both p53 and p63/73 proteins in

443 molluscs (111) we found five MtrBTN2-specific SNVs. The variants T177C and C816T are
444 synonymous substitutions previously described by Vassilenko et al. (2010) (112) in *M.*
445 *trossulus* affected by haemic neoplasia. The variant A1264T is a missense substitution, newly
446 described here, but is not located in a domain. *M. trossulus p53* is characterized by a
447 deletion of 6 nucleotides if compared to *M. edulis p53* (113). We found this deletion in 100%
448 of reads coming from MtrBTN2 samples. The frequencies of these MtrBTN2-specific SNVs
449 are reported in Supplementary Table S8, and in most cases, heterozygosity was observed,
450 with maintenance of expression of the *M. trossulus* alleles whereas host-specific (*M. edulis*)
451 alleles were predominantly absent (Supplementary Files S1). Indeed, the real impact of these
452 variants on the cellular functions is difficult to assess and this MtrBTN2-specific variant set
453 could include both somatic variants and ancestral germline polymorphisms.

454 **3. Discussion**

455 In their survey of bilaterian animals, Aktipis et al. (114) have revealed that all or almost all
456 bilaterians are susceptible to cancer. Molluscs (Lophotrochozoa) are no exception (20,22).
457 Although a high degree of genotypic and phenotypic diversity has been described among
458 mammalian cancers (115), molecular studies, mainly in humans and mice, have identified
459 common drivers of the cancer process (25,26). Still, oncogenic processes in invertebrate
460 phyla have been poorly studied, especially at the molecular level. Even if expression levels
461 are not the only factor determining the effect of genes on phenotypes, our present
462 transcriptomic data suggest that the most common alterations associated with oncogenesis
463 found in mammals are present in MtrBTN2 cells. Six of the major oncogenic signalling
464 pathways are likely to be altered. One of the most fundamental traits of cancer cells is their
465 high proliferation rate. We have previously found a mean doubling time of ~ 3 days for

466 MtrBTN2 (24). The present study provides information on the molecular bases of the high
467 proliferation of MtrBTN2. The overexpression of these proliferation-related pathways is
468 probably even more evident because we have used specialised cells (haemocytes) as a
469 reference. At present, the issue of haematopoiesis in bivalves is poorly understood;
470 however, some circulating haemocytes are capable of dividing and expressing genes
471 involved in cell proliferation, particularly during the immune response (116). The
472 proliferative activity of MtrBTN2 appears to be supported by metabolic rewiring, as is often
473 seen in cancer cells (77). Indeed, we highlight a possible central role for the Pi3K-AKT-
474 mTORC1 pathway, which is a master regulator in coupling cell growth and amino-acid, lipid,
475 and nucleotide metabolisms that are frequently modified in many mammalian cancers of
476 diverse origin (117-119). Our data also suggest inhibition of apoptosis. Dysregulation of
477 receptors and pathways involved in cell-ECM interactions (120) was also highlighted here,
478 which is consistent with our knowledge of MtrBTN2 biology, pathogenesis (intra-host
479 disease progression) and its inter-host transmission. Indeed, MtrBTN2 cells are non-adherent
480 circulating cells that infiltrate tissues and breach physiological barriers such as epithelium
481 during the transmission process. MtrBTN2 cells are also characterised by a dysregulated
482 differentiation state often seen in aggressive cancers (85), as reflected by the aberrant
483 activation of some meiosis-related genes suggesting a soma-to-germline transition. Such
484 reactivation of meiosis-related gene expression has previously been observed in a variety of
485 non-germ cell cancers in humans (89). Finally, our transcriptomic data suggest a dampening
486 of host defenses in cancerous mussels, which is likely to facilitate disease progression and
487 pathological outcome. Indeed, evasion from immune destruction by silencing components of
488 the immune system is another hallmark of human cancers (121).

489 This transcriptomic analysis generated a large amount of information about gene expression
490 in MtrBTN2 cells and we identified candidate genes and pathways associated with their
491 cancerous state. These results represent a significant step forward in understanding this
492 disease at the cellular and molecular level, and lay the groundwork for future research. To
493 date, most of the molecular dysfunctions have not been studied in molluscs and may differ
494 to some extent from vertebrates. Priorities are to i) obtain well curated and annotated
495 reference genomes, ii) accurately characterise gene products, and iii) specifically describe
496 protein functions and relationships within and between pathways. To achieve these goals,
497 gene manipulation tools are strongly required, which is a major challenge in such non-model
498 organisms. However, a comparative oncology approach offers a unique and powerful
499 opportunity to learn more about the evolutionary mechanisms of cancers and metastatic
500 processes. Our model has several advantages. It is a naturally occurring cancer, it is
501 transmissible, ensuring the repeatability of analyses and long-term studies, and molluscs are
502 easy to handle under laboratory conditions with few ethical concerns.

503

504 **5. Materials and Methods**

505 **5.1 Mussel samplings and MtrBTN2 diagnostic**

506 We collected mussels in the English Channel and the Barents Sea in late 2019. Two hundred
507 *Mytilus edulis* were sampled in a farm located in Agon-Coutainville (49°0'44.784''N
508 1°35'55.643''O, Normandy, France) and immediately screened for cancer at the LABEO
509 laboratory (Caen, France). The presence of circulating cancer cells and the stage of the
510 disease were initially diagnosed by cytological examination of haemolymph samples (19). We
511 found 14 positive individuals among which three were at an advanced stage of the disease

512 (>95% of circulating cells were cancer cells). After anaesthesia (122), we drew a maximum
513 volume of haemolymph from the adductor muscle of these three cancerous mussels, as well
514 as from three mussels diagnosed as MtrBTN2-free. The haemolymph was placed individually
515 in RNase-free microtubes kept on ice and centrifuged at 800 x g for 10 min at 4°C. The
516 pelleted cells were immediately resuspended in TRIzol® (invitrogen) and stored at -80°C until
517 RNA extraction. As non-transmissible circulating cancers also exist in mussels (20), we
518 confirmed the MtrBTN2 diagnosis by two qPCRs, one specific for *M. trossulus* (and MtrBTN)
519 and targeting the nuclear marker EF1α (24) and one specific for the MtrBTN2 lineage
520 targeting mtCR (16). Briefly, a piece of mantle and gills was fixed in 96% ethanol and used for
521 DNA extraction using the Nucleomag® 96 Tissue kit (MachereyNagel). We carried out both
522 qPCRs on cancerous and cancer-free *M. edulis* using the sensiFAST™SyBR® No-ROX Kit
523 (Bioline) and the LightCycler® 480 Real-Time PCR (Roche Life Science) system. A transversal
524 section (including gills, digestive gland, mantle, and gonad) of each individual was fixed in
525 Davidson for 48h and embedded it paraffin (RHEM facility). Sections were cut at 3 µm-thin
526 and stained with haematoxylin and eosin to confirm the health status of each mussel and to
527 exclude the presence of other pathologies.
528 Three *M. trossulus* were collected in the wild in Mishukovo (69°02'39;34.3''N
529 33°01'39;45.9''E, Kola Bay, Russia, Barents Sea). The haemolymph was drawn as for *M. edulis*
530 individuals, the cells were pelleted and resuspended in 100 µL of RNA*later*™ (invitrogen) and
531 sent to our laboratory where they were frozen at -80°C until RNA extraction. We also
532 received a piece of mantle and gills from each individual fixed in 96% ethanol for genetic
533 screening and to exclude the presence of MtrBTN, and tissue sections in 70% ethanol after

534 48h of fixation in Davidson's solution. *M. trossulus* individuals were subjected to both qPCRs
535 and histological examination.

536 **5.2 Total RNA extraction, library preparation, and sequencing**

537 *M. trossulus* samples preserved in RNA*later*™ were centrifuged at 5000 x g for 10 min at 4°C,
538 the supernatant removed, and the cells were resuspended in 500 µL of TRIzol®. *M. edulis*
539 preserved directly in TRIzol® were thawed. Both sample types were incubated at room
540 temperature for 20 min under agitation to lyse the cells. RNA extraction was performed
541 using Direct-zol™ RNA MiniPrep according to the manufacturer's instructions (Zymo
542 Research). Quantification and integrity of total RNA were checked using a NanoDrop® ND-
543 1000 spectrophotometer (Thermo Scientific) and by capillary electrophoresis on a 2100
544 Bioanalyzer (Agilent).

545 Polyadenylated RNA-seq library construction and sequencing using Illumina® technology
546 were performed by the GENEWIZ® Company (Germany). Four hundred ng of total RNA at a
547 concentration of 10 ng/µL, and OD260/280 comprised between 1.85 and 2.21 were used for
548 library preparation. The NEBnext® Ultra™ II Directional RNA kit was used for the cDNA
549 library preparation and 9 cycles of enrichment PCR were run. Sequencing was performed on
550 Illumina® NovaSeq™, with a 150bp paired-end configuration, and a sequencing depth of
551 100M raw paired-end reads per sample.

552 **5.3 Additional samples**

553 Sixteen additional mussel individuals were used to validate our results. Ten mussels (5
554 MtrBTN2-affected and 5 MtrBTN2-free) were collected in a distinct location, at Le Croisic
555 (47°17'58.5"N 2°31'00.4"W, Loire-Atlantique, France) in February 2023 and used to validate

556 RNAseq-based quantification and differential expression between MtrBTN2 cells and normal
557 haemocytes by RT-qPCRs (see Section 5.5). Total RNA extraction was performed as
558 previously described (see Section 5.2). RNAseq reads from 6 mussels (3 MtrBTN2-affected
559 and 3 MtrBTN2-free) collected in Agon-Coutainville (49°0'44.784''N 1°35'55.643''O,
560 Normandy, France) in February 2021 for a separate study were used to confirm the
561 differential transcriptomic profile of MtrBTN2 (see Section 5.5) and to validate the MtrBTN2-
562 specific SNPs (see Section 5.6). Whole flesh from these 6 individuals was individually frozen
563 in liquid nitrogen and powdered by cryogenic grinding (Mixer Mill MM 400, Retsch). For RNA
564 extraction, mussel powder (10 mg) was homogenised in 500 µL of TRIzol by vortexing for
565 1 h at 4 °C. Prior to extraction, performed as described above, insoluble material was
566 removed by centrifugation at 12,000 × g for 10 min at 4 °C.

567 **5.4 De novo transcriptome assembly and functional annotation**

568 Raw reads were processed using RCorrector v1.04
569 (https://anaconda.org/bioconda/rcorrector/files?version=1.0.4) with default settings and -rf
570 configuration to correct sequencing errors (123). Uncorrectable reads were then removed
571 using the FilterUncorrectabledPEfastq tool
572 (https://github.com/harvardinformatics/TranscriptomeAssemblyTools/blob/master/FilterUn
573 correctabledPEfastq.py). The output reads were further processed for adapter removal and
574 trimming with TrimGalore! v0.6.4 (https://github.com/FelixKrueger/TrimGalore) with default
575 parameters and -q 28, --length 100. Ribosomal RNAs potentially still present after polyA
576 capture were removed by alignment against the SILVA Ribosomal database with Bowtie2
577 v2.4.1 (124). Read quality was assessed before and after read trimming using FastQC v0.11.9
578 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

579 In France, MtrBTN2 infects *M. edulis* hosts, but the lineage originated in a *M. trossulus*
580 founder host. These two species are closely related, hybridise when they come into contact,
581 either naturally or via human-induced introductions (125-127), and introgression between
582 them is observed. However, *M. trossulus* is not present in France and individuals of this
583 species have been sampled in a different environment (Barent Sea *vs* Channel Sea for
584 MtrBTN2). Therefore, to allow a differential gene expression analysis of the cancer cells, all
585 retained reads (from MtrBTN2 cells, *M. edulis* haemocytes and *M. trossulus* haemocytes)
586 were assembled into a pantranscriptome using Trinity v2.8.5 (128) with the default options.
587 TransDecoder v3.0.1 (https://github.com/TransDecoder/TransDecoder/wiki) was run on
588 these contigs to identify CDSs with a minimum length of 90 amino acids. Finally, CDS
589 sequence redundancy was reduced using CD-HIT-EST v4.8.1
590 (https://github.com/weizhongli/cdhit/wiki) with the following options: -n 6, -c 0.86, -G 0, -aL
591 1, -aS 1. We assessed the quality of the assembly using several tests. The assembly was first
592 analysed using TrinityStats, and the final pantranscriptome completeness was estimated
593 using BUSCO v5.1.1 (129) against the conserved single-copy metazoan gene database (n =
594 978). Finally, filtered reads were mapped back to the filtered pantranscriptome to assess the
595 individual mapping rate with Bowtie2 v2.4.1 (124). To conclude that our pantranscriptome
596 strategy was reliable, we performed a Redundancy Analysis (RDA) on the normalised read
597 counts to analyse the impact of explanatory variables ("cell type": haemocyte/MtrBTN2
598 cells, "species": trossulus/edulis, "environment": Barents Sea/Channel Sea) on response
599 variables (gene expression), followed by an ANOVA-like permutation test (nperm=999,
600 model="full") (https://cran.r-project.org/web/packages/vegan/vegan.pdf) (supplementary
601 Figure S1). Both "cell type" and "environment" were retained as significant explanatory

602 variables ($p<0.05$). This confirms that the inclusion of healthy *M. edulis* mussels in the
603 analysis is necessary to subtract the environment effect (Channel Sea *vs* Barent Sea).

604 For functional annotation, the transcripts were searched against the Uniprot (Swissprot and
605 TrEMBL) protein reference database (130) using PLASTX v2.3.3 algorithm with an e-value
606 cutoff of 1.0E-3 (131). Domain prediction against the InterPro database (132) was carried out
607 with InterProScan v5.48-83.0. Both results were combined and we used the OmicsBox
608 program v2.0 (34,35) to assign GO terms to the annotated sequences with an e-value hit
609 filter of 1.10−3, an annotation cutoff of 55, a HSP-Hit coverage of 60%, and an evidence code
610 of 0.8. We searched for additional correspondences using EggNog v5.0 (133) by an orthology
611 analysis.

612 **5.5 Quantification of transcript expression**

613 To investigate differential expression between MtrBTN2 cells and *M. edulis-M. trossulus*
614 haemocytes, reads were first aligned to the pantranscriptome using SALMON v1.6 (134).
615 DETs were identified using DESeq2 v1.34 (36) with a FDR<0.05. Count normalisation was
616 performed using the RLE method implemented in the DESeq2 package (135). To confirm the
617 identified DETs, we ran the QuantEval pipeline (https://github.com/dn070017/QuantEval,
618 37) with the "contig mode" and DESeq2 on the quantification outputs. We compared the
619 quantification results by Spearman's correlation analysis and we retained the shared DETs
620 for the subsequent analyses.

621 To validate the quantifications and differential expression profiles obtained from our RNA-
622 seq data, we carried out two alternative analyses on additional samples (see Section 5.3).
623 First, we performed RT-qPCRs targeting 17 genes (*ADAMTS1, ALOX5, APC, BMP2, CCND1,*

624 *CTNNB1, CYP11A1, FBXW7, JAG1, MAD3, PIK3R1, PTEN, SAV1, SELP, SFRP1, SFRP3,* and
625 *THBS1*) selected for their contrasting expression levels according to the DESeq2 analysis and
626 their oncogenic relevance. Reverse transcription into cDNA was performed using M-MLV
627 Reverse Transcriptase (MMLV-RT, invitrogen) following the manufacturer's instructions. The
628 total RT-qPCR reaction volume was 1.5 µL and consisted of 1 µL of LightCycler 480 SYBR
629 Green I Master Mix (Roche) containing 0.5 µM of PCR primer (Eurogenetec) and 0.5 µl of
630 cDNA (1/10 dilution). The amplification efficiency of each primer pair (supplementary Table
631 S7) was validated by serial dilution of a pool of all cDNAs. Relative expression was calculated
632 as the threshold cycle (Cq) values of selected genes minus the mean of the measured
633 threshold cycle (Cq) values of three constitutively expressed genes (*GAPDH, RPS28,* and
634 *UBE2D3*) (supplementary Table S8). We compared the Log2FC values obtained from the two
635 quantification methods. Log2FCs for RT-qPCRs were calculated as follow:
636 FC= efficiency of the qPCR ^| mean of Ct of the 5 MtrBTN2-affected individuals - mean of Ct
637 of the 5 MtrBTN2-free individuals|
638 We also performed Spearman's correlation analysis between normalised expression
639 obtained from RNA-seq data (RLE values) and normalised Ct obtained from RT-qPCR.
640 Second, we performed DESeq2 analysis on complementary RNA-seq data obtained from the
641 whole flesh of 6 additional samples from a separate study (see Section 5.3) to confirm
642 transcripts specific to the malignant state.

643 **5.6 Biological interpretation of gene expression profiles**

644 In the context of a non-model species, we used three different approaches to interpret the
645 biological relevance of DETS.

646 ***Enrichment analysis***

647 We performed a GO term enrichment analysis focusing on biological processes with the
648 GO_MWU tool (https://github.com/z0on/GO_MWU) using adaptive clustering and a rank-
649 based statistical test (Mann–Whitney U-test). We used the following parameters for
650 adaptive clustering: largest = 0.5; smallest = 10; clusterCutHeight = 0.5. We considered
651 both the level of expression and the significance of the differential expression: we assigned
652 the log2 fold change value to transcripts that were significantly differentially expressed
653 (adjusted p < 0.05), while assigning zero to the others. We considered as enriched a
654 biological process with a FDR < 1%. To display the results synthetically, we used the
655 Enrichment Map v3.3.3 tool (136) in Cytoscape v3.9.1 (137). The intensity of the enrichment
656 was evidenced in the network and was calculated as follows: i) for the processes enriched
657 with over-expressed transcripts, "number of transcripts significantly over-expressed in the
658 process/total number of transcripts in the process"; (ii) for the processes enriched with
659 under-expressed transcripts, "−1 × (number of transcripts significantly downregulated in
660 the process/total number of transcripts in the process)".

661 ***Hub and top expressed gene identification***

662 The top 20 genes in terms of connectivity ranking in the PPI network were selected as Hub
663 genes. We used the Search Tool for the Retrieval of Interacting Genes (STRING), a biological
664 database designed to predict PPI networks (138). The results were visualised in Cytoscape
665 v3.9.1 using NetworkAnalyzer v4.4.8 visualisation software (139), which can construct
666 comprehensive models of biological interactions. Isolated and partially connected nodes
667 were not included.

668 To identify marker genes of the cancerous condition, we considered both differential
669 expression (Log2FC) and expression level (RLE) values to build a plot graph. We defined
670 arbitrarily thresholds (log2FC values > |5|and RLE > 500) to highlight the most discriminating
671 genes between the two conditions (cancerous/healthy circulating cells).

672 ***Targeted analyses***

673 We performed a targeted analysis focusing on genes and pathways that have been identified
674 by Sanchez-Vega et al. (26) as being altered at high frequency across many different human
675 cancer types. We searched for the presence of a differential expression of these genes in
676 MtrBTN2 cells or of MtrBTN2-specific alleles (i.e. present only in MtrBTN2-affected mussels,
677 and present in all MtrBTN2-affected mussels). We also identified *M. edulis* and *M. trossulus*-
678 specific alleles. Allele visualisation and calculation of variant frequencies were carried out
679 using the IGV v2.12.3 (140) and iVar (141) tools on BAM files obtained after read alignment
680 to the transcript sequences with Bowtie2. v2.4.1 (124). The minimum quality score threshold
681 to count a base and the minimum frequency threshold were set to 20 and 0.02, respectively,
682 with a minimum base coverage of 50 reads. MtrBTN2-specific alleles were searched in CAN1-
683 3 individuals and then confirmed in additional CAN4-6 individuals (see Section 5.3).

684

685 **References**

686 1. Ostrander EA, Davis BW, Ostrander GK. Transmissible tumors: breaking the cancer
687 paradigm. Trends Genet. 2016;32(1): 1-15. doi: 10.1016/j.tig.2015.10.001.

688 2. Dujon AM, Gatenby RA, Bramwell G, MacDonald N, Dohrmann E, Raven N, et al.
689 Transmissible cancers in an evolutionary perspective. iScience. 2020;23(7): 101269. doi:
690 10.1016/j.isci.2020.101269.
691 3. Nowinsky M. Zurfrageueber die Impfung der krebsigengeschwuelste. Zentralbl Med
692 Wissensch. 1876;14: 790-791.
693 4. Murgia C, Pritchard JK, Kim SY, Fassati A, Weiss RA. Clonal origin and evolution of a
694 transmissible cancer. Cell. 2006;126(3): 477-487. doi: 10.1016/j.cell.2006.05.051.
695 5. Baez-Ortega A, Gori K, Strakova A, Allen JL, Allum KM, Bansse-Issa L, Bhutia TN, Bisson JL,
696 Briceño C, Castillo Domracheva A, Corrigan AM, Cran HR, Crawford JT, Davis E, de Castro
697 KF, B de Nardi A, de Vos AP, Delgadillo Keenan L, Donelan EM, Espinoza Huerta AR,
698 Faramade IA, Fazil M, Fotopoulou E, Fruean SN, Gallardo-Arrieta F, Glebova O, Gouletsou
699 PG, Häfelin Manrique RF, Henriques JJGP, Horta RS, Ignatenko N, Kane Y, King C, Koenig
700 D, Krupa A, Kruzeniski SJ, Kwon YM, Lanza-Perea M, Lazyan M, Lopez Quintana AM,
701 Losfelt T, Marino G, Martínez Castañeda S, Martínez-López MF, Meyer M, Migneco EJ,
702 Nakanwagi B, Neal KB, Neunzig W, Ní Leathlobhair M, Nixon SJ, Ortega-Pacheco A,
703 Pedraza-Ordoñez F, Peleteiro MC, Polak K, Pye RJ, Reece JF, Rojas Gutierrez J, Sadia H,
704 Schmeling SK, Shamanova O, Sherlock AG, Stammnitz M, Steenland-Smit AE, Svitich A,
705 Tapia Martínez LJ, Thoya Ngoka I, Torres CG, Tudor EM, van der Wel MG, Vițălaru BA,
706 Vural SA, Walkinton O, Wang J, Wehrle-Martinez AS, Widdowson SAE, Stratton MR,
707 Alexandrov LB, Martincorena I, Murchison EP. Somatic evolution and global expansion of
708 an ancient transmissible cancer lineage. Science. 2019;365(6452): eaau9923. doi:
709 10.1126/science.aau9923.

710 6. Pye RJ, Pemberton D, Tovar C, Tubio JM, Dun KA, Fox S, Darby J, Hayes D, Knowles GW,
711 Kreiss A, Siddle HV, Swift K, Lyons AB, Murchison EP, Woods GM. A second transmissible
712 cancer in Tasmanian devils. Proc Natl Acad Sci U S A. 2016;113(2): 374-379. Doi:
713 10.1073/pnas.1519691113.
714 7. Siddle HV, Kreiss A, Eldridge MD, Noonan E, Clarke CJ, Pyecroft S, et al. Transmission of a
715 fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened
716 carnivorous marsupial. Proc Natl Acad Sci USA. 2007;104(41): 16221-16226. doi:
717 10.1073/pnas.0704580104.
718 8. Hamede RK, Bashford J, McCallum H, Jones M. Contact networks in a wild Tasmanian
719 devil (*Sarcophilus harrisii*) population: using social network analysis to reveal seasonal
720 variability in social behaviour and its implications for transmission of devil facial tumour
721 disease. Ecol Lett. 2009;12(11): 1147-1157. doi: 10.1111/j.1461-0248.2009.01370.x.
722 9. Hamede RK, McCallum H, Jones M. Biting injuries and transmission of Tasmanian devil
723 facial tumour disease. J Anim Ecol. 2013;82(1): 182-190. doi: 10.1111/j.1365-
724 2656.2012.02025.x.
725 10. Hawkins CE, Baars C, Hesterman H, Hocking GJ, Jones ME, Lazenby B, et al. Emerging
726 disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus*
727 *harrisii*. Biol Conserv. 2006;131(2): 307-324. doi: 10.1016/j.biocon.2006.04.010.
728 11. Lachish S, Jones M, McCallum H. The impact of disease on the survival and population
729 growth rate of the Tasmanian devil. J Anim Ecol. 2007;76(5): 926-936. doi:
730 10.1111/j.1365-2656.2007.01272.x.

731 12. McCallum H, Jones M, Hawkins C, Hamede R, Lachish S, Sinn DL, Beeton N, Lazenby B.
732 Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-
733 induced extinction. Ecology. 2009;90(12): 3379-3392. doi: 10.1890/08-1763.1.
734 13. Lazenby BT, Tobler MW, Brown WE, Hawkins CE, Hocking GJ, Hume F, et al. Density
735 trends and demographic signals uncover the long-term impact of transmissible cancer in
736 Tasmanian devils. J Appl Ecol. 2018;55(3): 1368-1379. doi: 10.1111/1365-2664.13088.
737 14. Metzger MJ, Reinisch C, Sherry J, Goff SP. Horizontal transmission of clonal cancer cells
738 causes leukemia in soft-shell clams. Cell. 2015;161(2): 255-263. doi:
739 10.1016/j.cell.2015.02.042.
740 15. Metzger MJ, Villalba A, Carballal MJ, Iglesias D, Sherry J, Reinisch C, et al. Widespread
741 transmission of independent cancer lineages within multiple bivalve species. Nature.
742 2016;534(7609): 705-709. doi: 10.1038/nature18599.
743 16. Yonemitsu MA, Giersch RM, Polo-Prieto M, Hammel M, Simon A, Cremonte F, et al. A
744 single clonal lineage of transmissible cancer identified in two marine mussel species in
745 South America and Europe. elife. 2019;8: e47788. doi: 10.7554/eLife.47788.
746 17. Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al.
747 Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion
748 between clam species in the Seas of Southern Europe. eLife, 2022;11: e66946.doi:
749 10.7554/eLife.66946
750 18. Riquet F, Simon A, Bierne N. Weird genotypes? Don't discard them, transmissible cancer
751 could be an explanation. Evol Appl. 2017;10(2): 140-145. doi: 10.1111/eva.12439.
752 19. Burioli EAV, Trancart S, Simon A, Bernard I, Charles M, Oden E, Bierne N, Houssin M.
753 Implementation of various approaches to study the prevalence, incidence and

754 progression of disseminated neoplasia in mussel stocks. J Invertebr Pathol. 2019;168:
755 107271. doi: 10.1016/j.jip.2019.107271.
756 20. Hammel M, Simon A, Arbiol C, Villalba A, Burioli EAV, Pépin JF, et al. Prevalence and
757 polymorphism of a mussel transmissible cancer in Europe. Mol Ecol. 2022;31(3): 736-
758 751. doi: 10.1111/mec.16052.
759 21. Skazina M, Odintsova N, Maiorova M, Ivanova A, Väinölä R, Strelkov P. First description
760 of a widespread *Mytilus trossulus*-derived bivalve transmissible cancer lineage in M.
761 trossulus itself. Sci Rep. 2021;11(1): 5809. doi: 10.1038/s41598-021-85098-5.
762 22. Carballal MJ, Barber BJ, Iglesias D, Villalba A. Neoplastic diseases of marine bivalves. J
763 Invertebr Pathol. 2015;131: 83-106. doi: 10.1016/J.JIP.2015.06.004.
764 23. Benadelmouna A, Saunier A, Ledu C, Travers MA, Morga B. Genomic abnormalities
765 affecting mussels (*Mytilus edulis-galloprovincialis*) in France are related to ongoing
766 neoplastic processes, evidenced by dual flow cytometry and cell monolayer analyses. J
767 Invertebr Pathol. 2018;157: 45-52. doi: 10.1016/j.jip.2018.08.003.
768 24. Burioli EAV, Hammel M, Bierne N, Thomas F, Houssin M, Destoumieux-Garzón D,
769 Charrière GM. Traits of a mussel transmissible cancer are reminiscent of a parasitic life
770 style. Sci Rep. 2021;11(1): 24110. doi: 10.1038/s41598-021-03598-w.
771 25. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1): 31-46. doi:
772 10.1158/2159-8290.CD-21-1059.
773 26. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling
774 pathways in the cancer genome atlas. Cell. 2018;173(2): 321-337.e10. doi:
775 10.1016/j.cell.2018.03.035.

776 27. Giersch RM, Hart SFM, Reddy SG, Yonemitsu MA, Orellana Rosales MJ, Korn M, Geleta
777 BM, Countway PD, Fernández Robledo JA, Metzger MJ. Survival and Detection of Bivalve
778 Transmissible Neoplasia from the Soft-Shell Clam Mya arenaria (MarBTN) in Seawater.
779 Pathogens. 2022;11(3):283. doi: 10.3390/pathogens11030283.
780 28. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular
781 signatures become relevant to patient care? Nat Rev Cancer. 2007;7(7): 545-553. doi:
782 10.1038/nrc2173.
783 29. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat
784 Rev Genet. 2009;10(1): 57-63. doi: 10.1038/nrg2484.
785 30. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A
786 survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17: 13. doi:
787 10.1186/s13059-016-0881-8. Erratum in: Genome Biol. 2016;17(1): 181.
788 31. Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet.
789 2011;12(10): 671-682. doi: 10.1038/nrg3068.
790 32. Bruzos AL, Santamarina M, García-Souto D, Díaz S, Rocha S, Zamora J, Lee Y, Viña-Feás A,
791 Quail MA, Otero I, Pequeño-Valtierra A, Temes J, Rodriguez-Castro J, Villanueva A, Costas
792 D, Rodríguez R, Prieto T, Tomás L, Alvariño P, Alonso J, Cao A, Iglesias D, Carballal MJ,
793 Amaral AM, Balseiro P, Calado R, El Khalfi B, Izagirre U, de Montaudouin X, Pade NG,
794 Probert I, Ricardo F, Ruiz P, Skazina M, Smolarz K, Pasantes JJ, Villalba A, Ning Z, Ju YS,
795 Posada D, Demeulemeester J, Baez-Ortega A, Tubio JMC. The evolution of two
796 transmissible leukaemias colonizing the coasts of Europe. BioRxiv. 2022.08.06.503021.
797 doi: https://doi.org/10.1101/2022.08.06.503021.

798 33. Payá-Milans M, Olmstead JW, Nunez G, Rinehart TA, Staton M. Comprehensive
799 evaluation of RNA-seq analysis pipelines in diploid and polyploid species. Gigascience.
800 2018;7(12):giy132. doi: 10.1093/gigascience/giy132.
801 34. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal
802 tool for annotation, visualisation and analysis in functional genomics research.
803 Bioinformatics. 2005;21: 3674-3676. doi: 10.1093/bioinformatics/bti610.
804 35. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-
805 throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids
806 res. 2008;36(10): 3420-3435. doi: 10.1093/nar/gkn176.
807 36. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
808 RNA-seq data with DESeq2. Genome Biol. 2014;15(12): 550. Doi: 10.1186/s13059-014-
809 0550-8.
810 37. Hsieh PH, Oyang YJ, Chen CY. Effect of de novo transcriptome assembly on transcript
811 quantification. Sci Rep. 2019;9(1):8304. doi: 10.1038/s41598-019-44499-3.
812 38. Hu J, Zhang Z, Shen WJ, Azhar S. Cellular cholesterol delivery, intracellular processing and
813 utilization for biosynthesis of steroid hormones. Nutr Metab (Lond). 2010;7: 47. doi:
814 10.1186/1743-7075-7-47.
815 39. Murphy-Ullrich JE, Poczatek M. Activation of latent TGF-β by thrombospondin-1:
816 mechanisms and physiology. Cytokine Growth Factor Rev. 2000;11(1-2): 59-69. doi:
817 10.1016/S1359-6101(99)00029-5.
818 40. Letterio JJ, Roberts AB. Regulation of immune responses by TGF-β. Annu Rev Immunol.
819 1998;16(1): 137-161. doi: 10.1146/annurev.immunol.16.1.137.

820 41. Ashcroft GS. Bidirectional regulation of macrophage function by TGF-β. Microbes Infect.
821 1999;1(15): 1275-1282. doi: 10.1016/S1286-4579(99)00257-9.
822 42. Läubli H, Borsig L. Selectins promote tumor metastasis. Semin Cancer Biol. 2010;20(3):
823 169-177. doi: 10.1016/j.semcancer.2010.04.005.
824 43. Fabricius HÅ, Starzonek S, Lange T. The role of platelet cell surface P-Selectin for the
825 direct platelet-tumor cell contact during metastasis formation in human tumors. Front
826 Oncol. 2021;11: 642761. doi: 10.3389/fonc.2021.642761.
827 44. Anderson KM, Seed T, Vos M, Mulshine J, Meng J, Alrefai W, Ou D, Harris JE. 5-
828 Lipoxygenase inhibitors reduce PC-3 cell proliferation and initiate nonnecrotic cell death.
829 Prostate. 1998;37(3): 161-173. doi: 10.1002/(sici)1097-0045(19981101)37:3<161::aid-
830 pros5>3.0.co;2-d.
831 45. Bishayee K, Khuda-Bukhsh AR. 5-Lipoxygenase antagonist therapy: a new approach
832 towards targeted cancer chemotherapy. Acta Biochim Biophys Sin. 2013;45(9): 709-719.
833 doi: 10.1093/abbs/gmt064.
834 46. Lee JH, Lee GT, Woo SH, Ha YS, Kwon SJ, Kim WJ, Kim IY. BMP-6 in renal cell carcinoma
835 promotes tumor proliferation through IL-10-dependent M2 polarization of tumor-
836 associated macrophages. Cancer Res. 2013;73(12): 3604-3614. doi: 10.1158/0008-
837 5472.CAN-12-4563.
838 47. Redondo-García S, Peris-Torres C, Caracuel-Peramos R, Rodríguez-Manzaneque JC.
839 ADAMTS proteases and the tumor immune microenvironment: Lessons from substrates
840 and pathologies. Matrix Biol Plus. 2020;9: 100054. doi: 10.1016/j.mbplus.2020.100054.
841 48. Rucci N, Sanità P, Angelucci A. Roles of metalloproteases in metastatic niche. Curr Mol
842 Med. 2011;11: 609-622. doi: 10.2174/156652411797536705.

843 49. Huang T, Sun L, Yuan X, Qiu H. Thrombospondin-1 is a multifaceted player in tumor
844 progression. Oncotarget. 2017;8(48): 84546-84558. doi: 10.18632/oncotarget.19165.
845 50. Bach DH, Park HJ, Lee SK. The dual role of bone morphogenetic proteins in cancer. Mol
846 Ther Oncolytics. 2017;8: 1-13. doi: 10.1016/j.omto.2017.10.002.
847 51. Yee KO, Connolly CM, Duquette M, Kazerounian S, Washington R, Lawler J. The effect of
848 thrombospondin-1 on breast cancer metastasis. Breast Cancer Res Treat. 2009;114(1):
849 85-96. doi: 10.1007/s10549-008-9992-6.
850 52. Nucera C, Porrello A, Antonello ZA, Mekel M, Nehs MA, Giordano TJ, et al. B-Raf(V600E)
851 and thrombospondin-1 promote thyroid cancer progression. Proc Natl Acad Sci USA.
852 2010;107(23): 10649-10654. doi: 10.1073/pnas.1004934107.
853 53. Borsotti P, Ghilardi C, Ostano P, Silini A, Dossi R, Pinessi D., et al. Thrombospondin-1 is
854 part of a Slug-independent motility and metastatic program in cutaneous melanoma, in
855 association with VEGFR-1 and FGF-2. Pigment Cell Melanoma Res. 2015;28: 73-81. doi:
856 10.1111/pcmr.12319.
857 54. Huang T, Wang L, Liu D, Li P, Xiong H, Zhuang L, Sun L, Yuan X, Qiu H. FGF7/FGFR2 signal
858 promotes invasion and migration in human gastric cancer through upregulation of
859 thrombospondin-1. Int J Oncol. 2017;50(5): 1501-1512. doi: 10.3892/ijo.2017.3927.
860 55. Iwamura T, Caffrey TC, Kitamura N, Yamanari H, Setoguchi T, Hollingsworth MA. P-
861 selectin expression in a metastatic pancreatic tumor cell line (SUIT-2). Cancer Res.
862 1997;57(6): 1206-1212.
863 56. Ding XZ, Kuszynski CA, El-Metwally TH, Adrian TE. Lipoxygenase inhibition induced
864 apoptosis, morphological changes, and carbonic anhydrase expression in human

865 pancreatic cancer cells. Biochem Biophys Res Commun. 1999;266(2): 392-399. doi:
866 10.1006/bbrc.1999.1824.
867 57. Gupta S, Srivastava M, Ahmad N, Sakamoto K, Bostwick DG, Mukhtar H. Lipoxygenase-5
868 is overexpressed in prostate adenocarcinoma. Cancer. 2001;91(4): 737-743. doi:
869 10.1002/1097-0142(20010215)91:4<737::aid-cncr1059>3.0.co;2-f.
870 58. Ohd JF, Nielsen CK, Campbell J, Landberg G, Löfberg H, Sjölander A. Expression of the
871 leukotriene D4 receptor CysLT1, COX-2, and other cell survival factors in colorectal
872 adenocarcinomas. Gastroenterology. 2003;124(1): 57-70. doi: 10.1053/gast.2003.50011.
873 59. Skovrlj B, Koehler SM, Anderson PA, Qureshi SA, Hecht AC, Iatridis JC, Cho SK. Association
874 Between BMP-2 and Carcinogenicity. Spine (Phila Pa 1976). 2015;40(23): 1862-1871. doi:
875 10.1097/BRS.0000000000001126.
876 60. Wu BX, Zeidan YH, Hannun YA. Downregulation of neutral ceramidase by gemcitabine:
877 Implications for cell cycle regulation. Biochim Biophys Acta. 2009;1791(8): 730-739. doi:
878 10.1016/j.bbalip.2009.03.012.
879 61. Loveday C, Tatton-Brown K, Clarke M, Westwood I, Renwick A, Ramsay E, et al.
880 Mutations in the PP2A regulatory subunit B family genes PPP2R5B, PPP2R5C and
881 PPP2R5D cause human overgrowth. Hum Mol Genet. 2015;24(17): 4775-4779. doi:
882 10.1093/hmg/ddv182.
883 62. Fu M, Wang C, Li Z, Sakamaki T, Pestell RG. Cyclin D1: normal and abnormal functions.
884 Endocrinology. 2004;145(12): 5439-5447. doi: 10.1210/en.2004-0959.
885 63. Pfeiffer V, Lingner J. Replication of telomeres and the regulation of telomerase. Cold
886 Spring Harbor Perspect Biol. 2013;5(5): a010405. doi: 10.1101/cshperspect.a010405.

887 64. Bell RJ, Rube HT, Xavier-Magalhães A, Costa BM, Mancini A, Song JS, Costello JF.
888 Understanding TERT Promoter Mutations: A Common Path to Immortality. Mol Cancer
889 Res. 2016;14(4): 315-323. doi: 10.1158/1541-7786.MCR-16-0003.
890 65. Mills GB, Schmandt R, McGill M, Amendola A, Hill M, Jacobs K, et al. Expression of TTK, a
891 novel human protein kinase, is associated with cell proliferation. J Biol Chem.
892 1992;267(22): 16000-16006. doi: 10.1016/S0021-9258(19)49633-6.
893 66. Tucker L, Zhang Q, Sheppard G, Lou P, Jiang F, McKeegan E, et al. Ectopic expression of
894 methionine aminopeptidase-2 causes cell transformation and stimulates proliferation.
895 Oncogene. 2008;27: 3967-3976. doi: 10.1038/onc.2008.14.
896 67. Nguyen JT, Ray C, Fox AL, Mendonça DB, Kim JK, Krebsbach PH. Mammalian EAK-7
897 activates alternative mTOR signaling to regulate cell proliferation and migration. Sci Adv.
898 2018;4(5):eaao5838. doi: 10.1126/sciadv.aao5838.
899 68. Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. Ragulator-Rag
900 complex targets mTORC1 to the lysosomal surface and is necessary for its activation by
901 amino acids. Cell. 2010;141(2): 290-303. doi: 10.1016/j.cell.2010.02.024.
902 69. Pastuszak-Lewandoska D, Czarnecka KH, Migdalska-Sęk M, Nawrot E, Domańska D,
903 Kiszałkiewicz J, et al. Decreased FAM107A expression in patients with non-small cell lung
904 cancer. Adv Exp Med Biol. 2015;852: 39-48. doi: 10.1007/5584_2014_109.
905 70. Saharinen J, Hyytiäinen M, Taipale J, Keski-Oja J. Latent transforming growth factor-beta
906 binding proteins (LTBPs)-structural extracellular matrix proteins for targeting TGF-beta
907 action. Cytokine Growth Factor Rev. 1999;10(2): 99-117. doi: 10.1016/s1359-
908 6101(99)00010-6.

909 71. Shimobayashi M, Hall M. Multiple amino acid sensing inputs to mTORC1. Cell Res.
910 2016;26: 7-20. doi: 10.1038/cr.2015.146.
911 72. Cargnello M, Tcherkezian J, Roux PP. The expanding role of mTOR in cancer cell growth
912 and proliferation. Mutagenesis. 2015;30(2):169-76. doi: 10.1093/mutage/geu045.
913 73. Ricoult SJ, Yecies JL, Ben-Sahra I, Manning BD. Oncogenic PI3K and K-Ras stimulate de
914 novo lipid synthesis through mTORC1 and SREBP. Oncogene. 2016;35(10):1250-60. doi:
915 10.1038/onc.2015.179.
916 74. Valvezan AJ, Turner M, Belaid A, Lam HC, Miller SK, McNamara MC, Baglini C, Housden
917 BE, Perrimon N, Kwiatkowski DJ, Asara JM, Henske EP, Manning BD. mTORC1 Couples
918 Nucleotide Synthesis to Nucleotide Demand Resulting in a Targetable Metabolic
919 Vulnerability. Cancer Cell. 2017;32(5):624-638.e5. doi: 10.1016/j.ccell.2017.09.013.
920 75. Ben-Sahra I, Hoxhaj G, Ricoult SJH, Asara JM, Manning BD. mTORC1 induces purine
921 synthesis through control of the mitochondrial tetrahydrofolate cycle. Science.
922 2016;351(6274):728-733. doi: 10.1126/science.aad0489.
923 76. Martínez-Reyes I, Chandel NS. Mitochondrial TCA cycle metabolites control physiology
924 and disease. Nat Commun. 2020;11: 102. doi: 10.1038/s41467-019-13668-3.
925 77. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The biology of cancer: metabolic
926 reprogramming fuels cell growth and proliferation. Cell Metab. 2008;7(1):11-20. doi:
927 10.1016/j.cmet.2007.10.002.
928 78. Nishimura K, Ishiai M, Horikawa K, Fukagawa T, Takata M, Takisawa H, Kanemaki MT.
929 Mcm8 and Mcm9 form a complex that functions in homologous recombination repair
930 induced by DNA interstrand crosslinks. Mol Cell. 2012;47(4): 511-522. doi:
931 10.1016/j.molcel.2012.05.047.

932 79. Hanada K, Budzowska M, Modesti M, Maas A, Wyman C, Essers J, Kanaar R. The
933 structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA
934 crosslinks into double-strands breaks. EMBO J. 2006;25(20): 4921-49. doi:
935 10.1038/sj.emboj.7601344.
936 80. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. Environ
937 Mol Mutagen. 2017;58(5): 235-263. doi: 10.1002/em.22087.
938 81. Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, Luise C, et al. Oncogene-
939 induced senescence is a DNA damage response triggered by DNA hyper-replication.
940 Nature. 2006;444(7119): 638-642. doi: 10.1038/nature05327.
941 82. Haber JE. Partners and pathways repairing a double-strand break. Trends Genet.
942 2000;16: 259-264. doi: 10.1016/s0168-9525(00)02022-9.
943 83. Kozakiewicz CP, Fraik AK, Patton AH, Ruiz-Aravena M, Hamilton DG, Hamede R, et al.
944 Spatial variation in gene expression of Tasmanian devil facial tumors despite minimal
945 host transcriptomic response to infection. BMC Genomics. 2021;22(1): 698. doi:
946 10.1186/s12864-021-07994-4.
947 84. Mizejewski GJ. Role of integrins in cancer: survey of expression patterns. Proc Soc Exp
948 Biol Med. 1999; 222(2):124-38. doi: 10.1046/j.1525-1373.1999.d01-122.x.
949 85. Jilkine A, Gutenkunst RN. Effect of dedifferentiation on time to mutation acquisition in
950 stem cell-driven cancers. PLoS Comput Biol. 2014;10(3):e1003481. doi:
951 10.1371/journal.pcbi.1003481.
952 86. Bocker T, Barusevicius A, Snowden T, Rasio D, Guerrette S, Robbins D, et al. hMSH5: A
953 human MutS homologue that forms a novel heterodimer with hMSH4 and is expressed
954 during spermatogenesis. Cancer res. 1999;59(4): 816-822.

955 87. Syrjänen JL, Pellegrini L, Davies OR. A molecular model for the role of SYCP3 in meiotic
956 chromosome organisation. eLife. 2014;3: e02963. doi: 10.7554/eLife.02963.
957 88. Adelman CA, Petrini JH. ZIP4H (TEX11) deficiency in the mouse impairs meiotic double
958 strand break repair and the regulation of crossing over. PLoS Genet. 2008;4(3):e1000042.
959 doi: 10.1371/journal.pgen.1000042.
960 89. McFarlane RJ, Wakeman JA. Meiosis-like Functions in Oncogenesis: A New View of
961 Cancer. Cancer Res. 2017;77(21):5712-5716. doi:10.1158/0008-5472.CAN-17-1535.
962 90. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens,
963 gametogenesis and cancer. Nat Rev Cancer. 2005;5(8):615-625.doi:10.1038/nrc1669.
964 91. Hubert F, Noel T, Roch P. A member of the arthropod defensing family from edible
965 Mediterranean mussels (*Mytilus galloprovincialis*). Eur J Biochem. 1996;240(1): 302-306.
966 doi: 10.1111/j.1432-1033.1996.0302h.x.
967 92. Mitta G, Hubert F, Noël T, Roch P. Myticin, a novel cysteine-rich antimicrobial peptide
968 isolated from haemocytes and plasma of the mussel *Mytilus galloprovincialis*. Eur J
969 Biochem. 1999;265(1): 71-78. doi: 10.1046/j.1432-1327.1999.00654.x.
970 93. Krasity BC, Troll JV, Weiss JP, McFall-Ngai MJ. LBP/BPI proteins and their relatives:
971 conservation over evolution and roles in mutualism. Biochem Soc T. 2011;39(4): 1039-
972 1044. doi: 10.1042/BST0391039.
973 94. Zhang L, Li L, Zhang G. A *Crassostrea gigas* Toll-like receptor and comparative analysis of
974 TLR pathway in invertebrates. Fish Shellfish Immunol. 2011;30(2): 653-660. doi:
975 10.1016/j.fsi.2010.12.023.

976 95. Garlatti V, Belloy N, Martin L, Lacroix M, Matsushita M, Endo Y, et al. Structural insights
977 into the innate immune recognition specificities of L- and H-ficolins. EMBO j. 2007;26(2):
978 623-633. doi: 10.1038/sj.emboj.7601500.
979 96. Schlosser A, Thomsen T, Moeller JB, Nielsen O, Tornøe I, Mollenhauer J, Moestrup SK,
980 Holmskov U. Characterization of FIBCD1 as an acetyl group-binding receptor that binds
981 chitin. J Immunol. 2009 ;183(6): 3800-3809. doi: 10.4049/jimmunol.0901526.
982 97. Dheilly NM, Duval D, Mouahid G, Emans R, Allienne JF, Galinier R, et al. A family of
983 variable immunoglobulin and lectin domain containing molecules in the snail
984 *Biomphalaria glabrata*. Dev Comp Immunol. 2015;48(1): 234-243. doi:
985 10.1016/j.dci.2014.10.009.
986 98. Trino S, Laurenzana I, Lamorte D, Calice G, De Stradis A, Santodirocco M, Sgambato A,
987 Caivano A, De Luca L. Acute myeloid leukemia cells functionally compromise
988 hematopoietic stem/progenitor cells inhibiting normal hematopoiesis through the release
989 of extracellular vesicles. Front Oncol. 2022;12: 824562. doi: 10.3389/fonc.2022.824562.
990 99. Kent ML, Elston RA, Wilkinson MT, Drum AS. Impaired defense mechanisms in bay
991 mussels, *Mytilus edulis*, with hemic neoplasia. J Invertebr Pathol. 1989;53(3): 378-386.
992 doi: 10.1016/0022-2011(89)90103-1.
993 100. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological
994 parameters of biological networks. Bioinformatics. 2008;24(2): 282-284. doi:
995 10.1093/bioinformatics/btm554.
996 101. Pankov R, Yamada KM. Fibronectin at a glance. J Cell Sci. 2002;115(20): 3861-3863.
997 doi: 10.1242/jcs.00059.

998 102. Cotteret S, Chernoff J. The evolutionary history of effectors downstream of Cdc42
999 and Rac. Genome Biol. 2002;3(2):REVIEWS0002. doi: 10.1186/gb-2002-3-2-reviews0002
1000 103. Porter AG, Jänicke RU. Emerging roles of caspase-3 in apoptosis. Cell Death Differ.
1001 1999;6(2): 99-104. doi: 10.1038/sj.cdd.4400476.
1002 104. Yu FX, Guan KL. The Hippo pathway: regulators and regulations. Genes Dev.
1003 2013;27(4): 355-371. doi: 10.1101/gad.210773.112.
1004 105. Carracedo A, Pandolfi P. The PTEN–PI3K pathway: of feedbacks and cross-talks.
1005 Oncogene. 2008;27: 5527-5541. doi: 10.1038/onc.2008.247.
1006 106. Cohen P, Frame S. The renaissance of GSK3. Nat Rev Mol Cell Biol. 2001;2: 769-776.
1007 doi: 10.1038/35096075.
1008 107. Amati B, Land H. Myc-Max-Mad: a transcription factor network controlling cell cycle
1009 progression, differentiation and death. Curr Opin Genet Dev. 1994;4(1): 102-108. doi:
1010 10.1016/0959-437x(94)90098-1.
1011 108. Dikovskaya D, Khoudoli G, Newton IP, Chadha GS, Klotz D, Visvanathan A, et al. The
1012 Adenomatous Polyposis Coli protein contributes to normal compaction of mitotic
1013 chromatin. PLOS One. 2012;7(6): e38102. doi: 10.1371/journal.pone.0038102.
1014 109. Lecarpentier Y, Schussler O, Hébert JL, Vallée A. Multiple targets of the canonical
1015 WNT/β-Catenin signaling in cancers. Front Oncol. 2019;9: 1248. doi:
1016 10.3389/fonc.2019.01248.
1017 110. Iso T, Kedes L, Hamamori Y. HES and HERP families: multiple effectors of the Notch
1018 signaling pathway. J Cell Physiol. 2003;194(3): 237-255. doi: 10.1002/jcp.10208.

1019 111. Kelley ML, Winge P, Heaney JD, Stephens RE, Farell JH, Beneden RJV, et al. Expression
1020 of homologues for p53 and p73 in the softshell clam (*Mya arenaria*), a naturally-occuring
1021 model for human cancer. Oncogene. 2001;20: 748-758. doi: 10.1038/sj.onc.1204144.
1022 112. Vassilenko EI, Muttray AF, Schulte PM, Baldwin SA. Variations in p53-like cDNA
1023 sequence are correlated with mussel haemic neoplasia: A potential molecular-level tool
1024 for biomonitoring. Mutat Res. 2010;701(2): 145-152. doi:
1025 10.1016/j.mrgentox.2010.06.001.
1026 113. Muttray AF, Cox RL, Reinisch CL, Baldwin SA. Identification of DeltaN isoform and
1027 polyadenylation site choice variants in molluscan p63/p73-like homologues. Mar
1028 Biotechnol (NY). 2007;9(2): 217-230. doi: 10.1007/s10126-006-6045-1.
1029 114. Aktipis CA, Boddy AM, Jansen G, et al. Cancer across the tree of life: cooperation and
1030 cheating in multicellularity. Philos Trans R Soc Lond B Biol Sci. 2015;370(1673):
1031 20140219. doi:10.1098/rstb.2014.0219.
1032 115. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al.
1033 Mutational heterogeneity in cancer and the search for new cancer-associated genes.
1034 Nature. 2013;499(7457): 214-218. doi: 10.1038/nature12213.
1035 116. Bachère E, Gueguen Y, Gonzalez M, de Lorgeril J, Garnier J, Romestand B. Insights
1036 into the anti-microbial defense of marine invertebrates: the penaeid shrimps and the
1037 oyster *Crassostrea gigas*. Immunol Rev. 2004;198:149-68. doi: 10.1111/j.0105-
1038 2896.2004.00115.x.
1039 117. Hoxhaj G, Manning BD. The PI3K-AKT network at the interface of oncogenic signalling
1040 and cancer metabolism. Nat Rev Cancer. 2020;20(2):74-88. doi: 10.1038/s41568-019-
1041 0216-7.

1042 118. Magaway C, Kim E, Jacinto E. Targeting mTOR and Metabolism in Cancer: Lessons and
1043 Innovations. Cells. 2019;8(12):1584. doi: 10.3390/cells8121584.
1044 119. Zou Z, Tao T, Li H, Zhu X. mTOR signaling pathway and mTOR inhibitors in cancer:
1045 progress and challenges. Cell Biosci. 2020;10:31. doi: 10.1186/s13578-020-00396-1.
1046 120. Hamidi H, Ivaska J. Every step of the way: integrins in cancer progression and
1047 metastasis. Nat Rev Cancer. 2018;18(9):533-548. doi: 10.1038/s41568-018-0038-z.
1048 121. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1): 31-46.
1049 doi: 10.1158/2159-8290.
1050 122. Suquet M, de Kermoysan G, Gonzalez Araya R, Queau I, Lebrun L, Le Souchu
1051 P, Mingant C. Anesthesia in Pacific oyster *Crassostrea gigas*. Aquat Living Resour.
1052 2009;22: 29-34. doi: 10.1051/alr/2009006.
1053 123. Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-
1054 seq reads. GigaScience. 2015;4(1): 48. doi: 10.1186/s13742-015-0089-y.
1055 124. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
1056 2012;9(4): 357-359. doi:10.1038/nmeth.1923.
1057 125. Fraïsse C, Belkhir K, Welch JJ, Bierne N. Local interspecies introgression is the main
1058 cause of extreme levels of intraspecific differentiation in mussels. Mol Ecol. 2016;25(1):
1059 269-286. doi: 10.1111/mec.13299.
1060 126. Popovic I, Matias AMA, Bierne N, Riginos C. Twin introductions by independent
1061 invader mussel lineages are both associated with recent admixture with a native
1062 congener in Australia. Evol Appl. 2020;13(3): 515-532. doi: 10.1111/eva.12857.

1063 127. Simon A, Arbiol C, Nielsen EE, Couteau J, Sussarellu R, Burgeot T, et al. Replicated
1064 anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels.
1065 Evol Appl. 2020;13(3): 575-599. doi: 10.1111/eva.12879.
1066 128. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length
1067 transcriptome assembly from RNA-Seq data without a reference genome. Nat
1068 Biotechnol. 2011;29(7): 644-652. doi: 10.1038/nbt.1883.
1069 129. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:
1070 assessing genome assembly and annotation completeness with single-copy orthologs.
1071 Bioinformatics. 2015;31: 3210-3212. doi: 10.1093/bioinformatics/btv351.
1072 130. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its
1073 supplement TrEMBL in 2000. Nucleic Acids Res. 2000;28: 45-48. doi:
1074 10.1093/nar/28.1.45.
1075 131. Nguyen VH, Lavenier D. PLAST: parallel local alignment search tool for database
1076 comparison. BMC Bioinf. 2009;10: 329. doi: 10.1186/1471-2105-10-329.
1077 132. Blum M, Chang H, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The
1078 InterPro protein families and domains database: 20 years on. Nucleic Acids Res.
1079 2021;49(D1): 344-354. doi: 10.1093/nar/gkaa977.. doi: 10.1093/nar/gkaa977.
1080 133. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al.
1081 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology
1082 resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019; 47(D1):
1083 309-314. doi: 10.1093/nar/gky1085.

1084 134. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-
1085 aware quantification of transcript expression. Nat Methods. 2017;14(4): 417-419. doi:
1086 10.1038/nmeth.4197.
1087 135. Anders S, Huber W. Differential expression analysis for sequence count data.
1088 Genome Biol. 2010;11(10):NR106. doi: 10.1186/gb-2010-11-10-r106.
1089 136. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based
1090 method for gene-set enrichment visualisation and interpretation. PLoS One.
1091 2010;5(11):e13984. doi: 10.1371/journal.pone.0013984.
1092 137. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a
1093 software environment for integrated models of biomolecular interaction networks.
1094 Genome Res. 2003;13(11): 2498-2504. doi: 10.1101/gr.1239303.
1095 138. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING
1096 database in 2021: customizable protein–protein networks, and functional
1097 characterization of user-uploaded gene/measurement sets. Nucleic Acids Res.
1098 2021;49(D1): 605-612. doi: 10.1093/nar/gkaa1074.
1099 139. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological
1100 parameters of biological networks. Bioinformatics. 2008;24(2):282-4. doi:
1101 10.1093/bioinformatics/btm554.
1102 140. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with
1103 the Integrative Genomics Viewer (IGV). Cancer Res. 2017;77(21): 31-34. doi:
1104 10.1158/0008-5472.CAN-17-0337.
1105 141. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, Tan AL,
1106 Paul LM, Brackney DE, Grewal S, Gurfield N, Van Rompay KKA, Isern S, Michael SF, Coffey

1107 LL, Loman NJ, Andersen KG. An amplicon-based sequencing framework for accurately
1108 measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20(1):8.
1109 doi: 10.1186/s13059-018-1618-7.

1110 **Acknowledgments**

1111 We acknowledge the Animal Health-Clinical Biology unit of the LABÉO laboratory (Caen,
1112 France), especially Maryline Houssin and Ludovic Petit, for their help during cytological
1113 screening and Professor Petr Strelkov from the Saint Petersburg State University for the
1114 sampling of *M. trossulus* individuals. We thank the "Réseau d'Histologie Expérimentale de
1115 Montpellier" - RHEM facility - supported by SIRIC Montpellier Cancer (Grant
1116 INCa_Inserm_DGOS_12553), the European regional development foundation and the
1117 Occitanian region (FEDER-FSE 2014-2020 Languedoc Roussillon), for processing our animal
1118 tissues, histology technics and expertise. This study is set within the framework of the
1119 'Laboratoires d'Excellence (LABEX)' Tulip (ANR- 10-LABX-41). **Funding:** this work was
1120 supported by the French National Agency for Research, TRANCAN project (ANR-18-CE35-
1121 0009) and the Montpellier Université d'Excellence (MUSE), BLUECANCER project. FT is
1122 supported by the MAVA Foundation. **Author contributions:** Conceptualization: EAVB, DDG,
1123 GM, GMC, EV, JVD; Investigation, Formal analysis: EAVB, MH; Methodology: EAVB, EV, JVD;
1124 Funding Acquisition: DDG, GMC, NB, FT; Original Draft preparation: EAVB;
1125 Reviewing&Editing: DDG, GMC, MH, EV, JVD, NB, FT, GM. **Competing interests:** The authors
1126 declare that they have no competing interests. **Data and materials availability:** Raw read
1127 sequences were deposited with links to BioProject accession number PRJNA749800 in the
1128 NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

1129