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

Erika A.V. Burioli, Maurine Hammel, Emmanuel Vignal, Jeremie Vidal-Dupiol, Guillaume Mitta, Frédéric Thomas, Nicolas Bierne, Delphine Destoumieux-Garzon, Guillaume Charrière

OPEN BIOLOGY

royalsocietypublishing.org/journal/rsob

Research

Cite this article: Burioli EAV, Hammel M, Vignal E, Vidal-Dupiol J, Mitta G, Thomas F, Bierne N, Destoumieux-Garzón D, Charrière GM. 2023 Transcriptomics of mussel transmissible cancer MtrBTN2 suggests accumulation of multiple cancer traits and oncogenic pathways shared among bilaterians. *Open Biol.* **13**: 230259.
https://doi.org/10.1098/rsob.230259

Received: 2 August 2023
Accepted: 12 September 2023

Subject Area:
cellular biology/molecular biology

Keywords:
transmissible cancer, mussel, transcriptomics, oncogenic pathways, MtrBTN2, *Mytilus*

Author for correspondence:
E. A. V. Burioli
e-mail: erikaastrid.burioli@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6873517.

THE ROYAL SOCIETY PUBLISHING

Transcriptomics of mussel transmissible cancer MtrBTN2 suggests accumulation of multiple cancer 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] and 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

EAVB, 0000-0003-2666-9258; MH, 0000-0001-9937-5678; EV, 0000-0002-2585-119X; JV-D, 0000-0002-0577-2953; GM, 0000-0003-1188-1467; FT, 0000-0003-2238-1978; NB, 0000-0003-1856-3197; DD-G, 0000-0002-8793-9138; GMC, 0000-0002-4796-1488

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 may 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.

1. Introduction

Transmissible cancers are rare and fascinating biological entities because they have evolved the ability to overcome the physical and immunological barriers of the host to become contagious and to spread between animals by direct transfer of cancer cells, thus behaving like parasites [1,2]. Among vertebrates, they have only been reported in two species so far, in dogs and in Tasmanian devils. The transmissibility of cancer cells was first demonstrated in the canine transmissible veneral tumour (CTVT) [3,4], which is thought to have originated more than 4000 years ago [5]. CTVT has since spread across continents, probably together with humans, through coitus and oral contact between dogs. Tasmanian devil facial tumour disease (DFTD) is the other well-known transmissible cancer that affects vertebrates. Two independent cancerous lineages, DFT1 and DFT2, have been described to date and have emerged less than fifty years ago [6]. Infection occurs via bites, a common behaviour in Tasmanian devil social interactions [7–9]. As a result, cancer transmission has devastated much of the devil population, with some local populations declining by as much as 80% in the geographical areas most affected by DFTD [10–13]. However, the highest number of transmissible cancer lineages identified to date

has been found in bivalve molluscs (8 different lineages) and these multiple lineages have been grouped under the term bivalve transmissible neoplasia (BTN) [14–17]. BTNs usually transmit between individuals of the same species, but some lineages have crossed the species barrier and are circulating in populations of related host species [15–17]. Indeed, one specific lineage of BTN, called *Mytilus trossulus* bivalve transmissible neoplasia lineage 2 (MtrBTN2), has emerged in a *M. trossulus* founder host and has since spread to other *Mytilus* species populations (*M. trossulus*, *M. edulis*, *M. galloprovincialis* and *M. chilensis*) across several continents—South America, Asia and Europe [15,16,18–21].

Transmissible cancer lineages must have emerged from a first neoplastic transformation in a founder host and then evolved specific phenotypes to become a new type of contagious etiologic agent. In fact, we have previously shown that conventional non-transmissible cancers also occur in *Mytilus* bivalves [20]. MtrBTN2 cells are found circulating in the haemolymph together with haemocytes, the mussel's immune cells. Like haemocytes, they are able to infiltrate connective tissues of various organs [19,22]. As the disease progresses, MtrBTN2 cells outgrow the host cells and progressively replace almost all the circulating cells present in the haemolymph and disseminate in the tissues. They have a characteristic morphology of rounded and basophilic cells with a high nucleus-to-cell ratio and they are polyploid [19]. Therefore, like other haemolymphatic cancers [20], they are easily diagnosed by histological/cytological observation or flow cytometry of haemolymph [19,23]. This initial diagnosis needs to be complemented by genetic analysis to distinguish MtrBTN2 from other transmissible cancers and conventional haemolymphatic cancers [16,24]. MtrBTN2 cells proliferate very quickly with a mean doubling time of approximately 3 days [24]. In addition, BTNs must be able to survive in the extra-host environment long enough to infect a new host [24,25]. We have shown that MtrBTN2 cells can survive at least for 3 days without mortality and up to 8 days in seawater [24]. However, the molecular pathways related to the malignancy of BTNs have not been characterized. Based on the phenotypic traits observed in MtrBTN2 cells, such as high proliferative activity, genomic abnormalities such as polyploidy, supermetastatic ability and extended cell survival capabilities, it is likely that multiple molecular pathways contribute to the malignancy and long-term persistence of this lineage.

Among the most common and most studied functional capabilities observed in human cancer cells are the sustaining of proliferative signalling, the evasion from growth suppressors, the resistance to cell death, the replicative immortality, the activation of invasion and metastasis, the reprogramming of cellular metabolism, and the avoidance of immune destruction [26]. The acquisition of these capabilities is ensured by the activation or inactivation of specific signalling pathways. Among these, Sanchez-Vega *et al*. [27] highlighted ten oncogenic signalling pathways that are most frequently altered in most human cancers (HIPPO, MYC, NOTCH, NRF2, PI3K, RTK/RAS, TGFβ, P53, WNT and cell cycle), and are mainly involved in promoting cell proliferation. These conclusions were drawn from extensive data on hundreds of human cancers obtained by transcriptomic, genomic and epigenomic analyses.

Measuring relative levels of gene expression has been a key approach to identify genes and biological pathways associated with the cancerous process and functional adaptations of cancers [28]. In recent years, RNA sequencing (RNA-seq) has emerged as a fast, cost-effective, and robust technology to address various biological questions [29,30]. Transcriptome analyses allow one to link cellular phenotypes to their molecular underpinnings. In the context of cancers, these links provide an opportunity to dissect the complexity of the cancer biological adaptations. For non-model organisms and in the absence of a suitable reference genome, as is the case for *M. trossulus*, RNA-seq is used to reconstruct and quantify de novo transcriptomes [31]. Differential gene expression analysis is then carried out to compare the effect of treatments or conditions on gene expression. To date, although several transmissible neoplasias have been described in marine bivalves, transcriptome-wide functional studies are still lacking. However, Bruzos *et al*. [32] have performed a comparative quantitative transcriptomic study between cancer cells and various healthy tissues in cockle BTNs, which supports their status as haemocyte-derived marine leukaemias.

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

2. Results

MtrBTN2 cells were collected from the haemolymph of *M. edulis* mussels reared in the English Channel (Normandy, France). Only individuals with more than 95% of circulating neoplastic cells in their haemolymph were used for MtrBTN2 cell collection. *Mytilus edulis* haemocytes were collected from individuals from the same geographical location but tested negative for MtrBTN2 by cytology and qPCRs targeting a sequence in the EF1 gene specific for the *M. trossulus* species and a MtrBTN2 specific sequence in the mitochondrial COI gene [16,24]. *Mytilus trossulus* haemocytes were collected from the haemolymph of wild *M. trossulus* found in the Barents Sea (Kola Bay, Russia), as this species is absent from the French coasts. All *M. trossulus* mussels were positive for the qPCR targeting the *M. trossulus* specific sequence, but negative by cytology and qPCR for MtrBTN2. We did not observe any other 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* (CAN1–3), MtrBTN2-negative *M. edulis* (EDU1–3) and MtrBTN2-negative *M. trossulus* (TRO1–3). All RNA-seq reads were used to assemble de novo a pantranscriptome from the 9 samples (CAN1–3, EDU1–3 and TRO1–3). The graphical representation of the experimental setup and bioinformatic analyses is shown in figure 1.

2.1. Pantranscriptome assembly and annotation

The raw assembly contained 2 149 788 transcripts. After filtering, we retained a total of 138 071 transcripts consisting of predicted coding sequences (CDS). The pantranscriptome

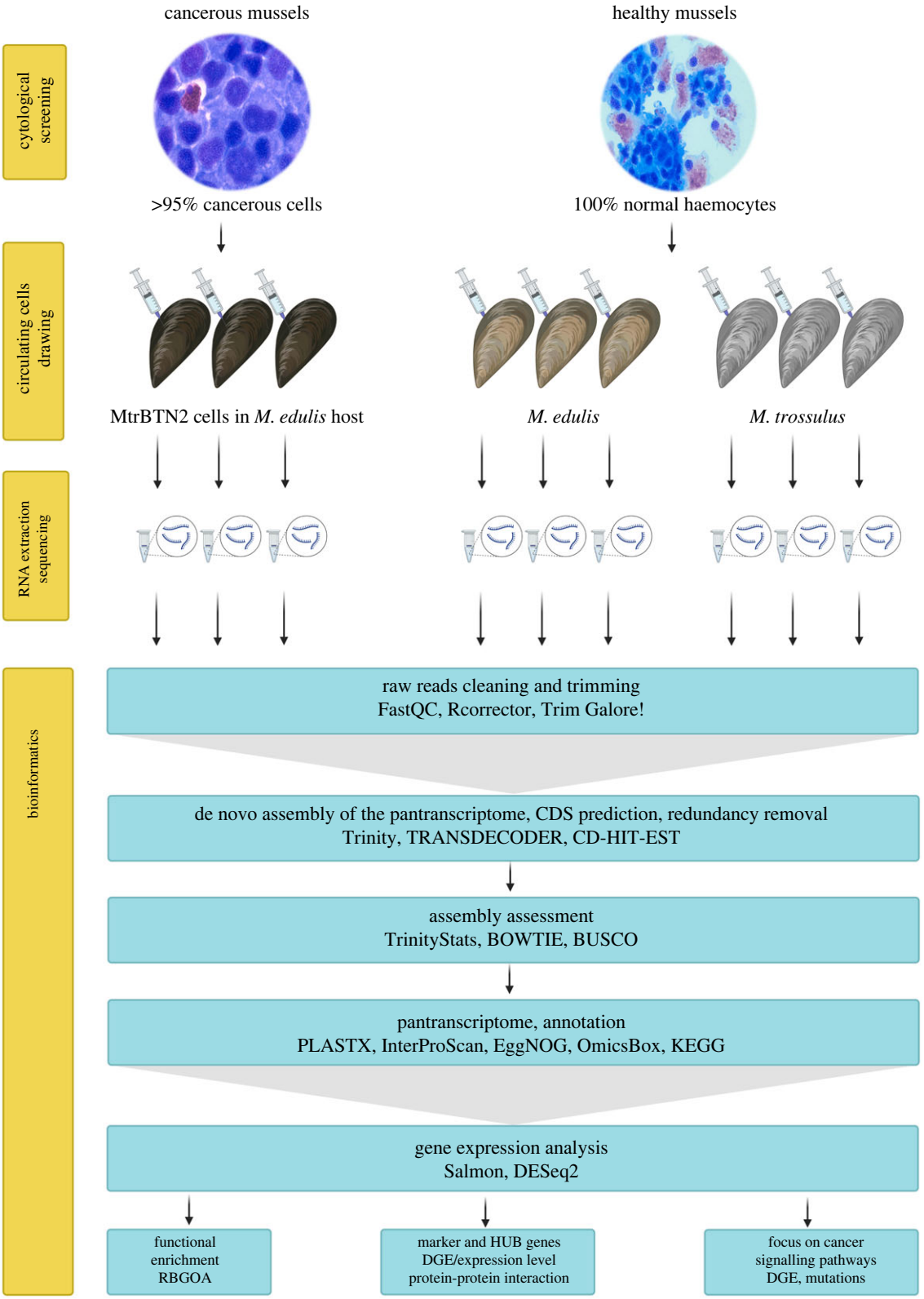

Figure 1. Graphical representation of the experimental design and RNA-Seq data analysis. DGE: differential gene expression analysis. Microphotographs of cancerous and healthy circulating cells are on the same scale.

completeness assessment revealed that 95.6% of the highly conserved single-copy metazoan genes ($n = 954$) were present in full length as a single (81.2%) or duplicated (14.4%) copies. The mean remapping rate was 53.38 ± 0.58% with negligible differences between samples (Kruskal–Wallis, $H(2) = 1.69$, $p = 0.43$), which is acceptable for a remapping on CDS [33]. Only 5826 transcripts showed no blast similarity or a conserved domain. The OmicsBox program v2.0 [34,35] assigned GO terms to 131 346 transcripts, but ultimately 41.39% (57 154) of transcript annotations were retained after computing a filtration step based on a minimum annotation score (electronic supplementary material, table S1). The assembly and annotation metrics are shown in table 1. Raw read sequences coming from RNA-seq were deposited in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) with links to BioProject accession number PRJNA749800. Individual SRA numbers are listed in electronic supplementary material, table S2.

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 (greater than 95% of cancer cells).

	cancerous individuals			healthy individuals					
	MtrBTN2			*Mytilus edulis*			*Mytilus trossulus*		
attributes	**CAN1**	**CAN2**	**CAN3**	**EDU1**	**EDU2**	**EDU3**	**TRO1**	**TRO2**	**TRO3**
clean read pairs $\times 10^6$	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					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	53.7	52.18	53.14
annotation statistics									
	transcripts with GO associated	131 346							
	transcripts annotated[a]	57 154							
	no-blast no. InterProScan hits found	5826							

[a]After applying the OmicsBox annotation rule (e-value hit filter of 1×10^{-3}, annotation cutoff of 55, HSP-Hit coverage of 60%).

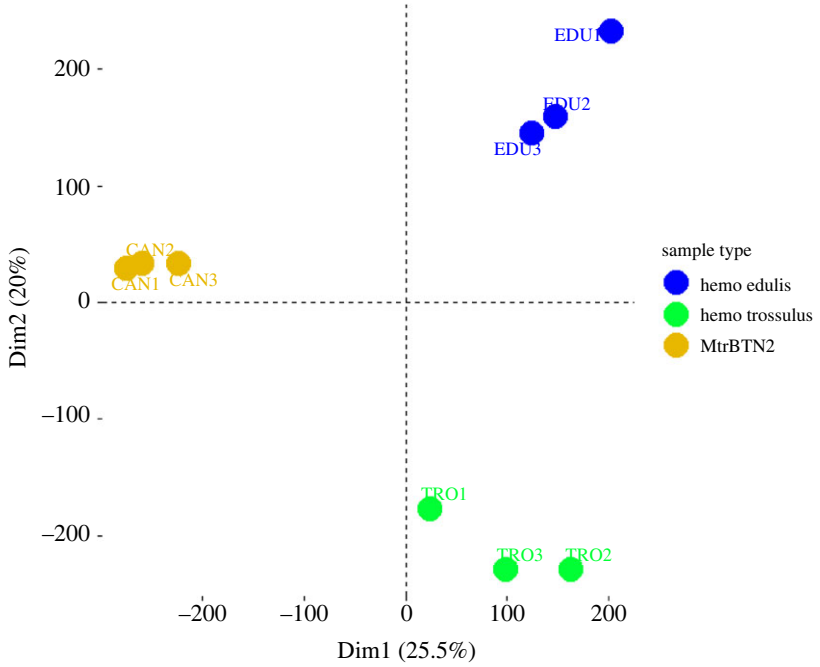

Figure 2. Principal component analysis (PCA) based on the normalized count matrix (RLE values). The first axis discriminates MtrBTN2 cells from haemocytes.

2.2. The specific transcriptomic profile of MtrBTN2 cells

Principal component analysis (PCA) based on the normalized count matrix (relative log expression (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.

We investigated differentially expressed transcripts in MtrBTN2 cells (CAN1–3) by comparing their abundance with *M. edulis* (EDU1–3) and *M. trossulus* haemocytes (TRO1–3) using DESeq2 v1.34 [36]. These comparative results were then validated by RT-qPCRs (see §§4.3 and 4.5) targeting 17 key-gene transcripts and confirmed in 10 additional haemolymph samples (CAN7–11, EDU7–11) (see electronic supplementary material, table S3, figures S1 and S2). A total of 4356/138 071 transcripts were significantly differentially expressed; 2632 were more expressed and 1724 were less expressed in MtrBTN2 cells than in haemocytes (see log2 fold change (Log2FC) and adjusted *p*-values (FDR) in electronic supplementary material, table S4). We further validated the significance of differential expression for 4254 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 (electronic supplementary material, tables S4 and S5, figure S3). Indeed, for 102 transcripts, the significance of the differential expression was not confirmed at the component level. A hierarchical clustering of differentially expressed transcripts (DETs) is shown on a log2 centred-heatmap (electronic supplementary material, figure S4). Although the results were less contrasted with RNA-seq data obtained from the whole flesh of another 6 additional individuals coming from a separate study (CAN4–6, EDU4–6) (see §§4.3 and 4.4; electronic supplementary material, table S3) than from haemolymph cells, the PCA of the transcript abundances also confirmed a differential clustering between MtrBTN2-affected and MtrBTN2-free mussels (electronic supplementary material, figure S5).

Among the most differentially expressed transcripts (with log2FC > 5), we identified six top-expressed (RLE > 500) transcripts of the genes *CYP11A1*, *THBS1*, *SELP*, *ALOX5*, *BMP2* and *ADAMTS1* (figure 3). These transcripts were also confirmed among the most differentially and top-expressed in the whole flesh of the additional 3 individuals (CAN4–6) (see §§4.3 and 4.5; electronic supplementary material, table S3, figure S6). Indeed, these genes can be considered as specific for the malignant state and have been described to be involved in human tumorigenesis and metastasis (table 2).

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

2.3. Dysregulated biological processes in MtrBTN2

To identify the key biological processes altered in the MtrBTN2 phenotype, we performed a GO_MWU rank-based

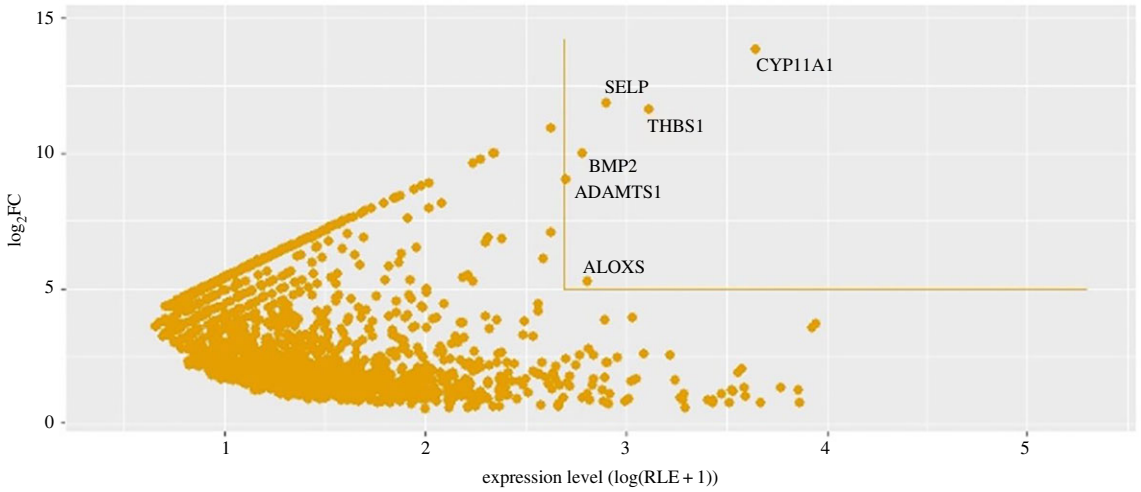

Figure 3. Scatter plot showing genes specific of the malignant state with $y = \log2FC$, $x = \log(RLE + 1)$. Lines delimit the areas above log2FC values > 5 and RLE > 500. Six transcripts of genes (*CYP11A1*, *THBS1*, *SELP*, *ALOX5*, *BMP2* and *ADAMTS1)* exceeded these values and can be considered as specific of the malignant state.

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

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

enrichment analysis on the DETs (for protein descriptions and gene IDs, see electronic supplementary material, table S4). A biological process is considered to be altered if DETs are significantly overrepresented (enriched) within the total set of transcripts related to that process. The complete list of transcripts clustered by Gene Ontology (GO) term is provided in the electronic supplementary material (table S6). The GO_MWU analysis revealed several biological functions that were differentially regulated in MtrBTN2 cells compared to normal haemocytes (figure 4).

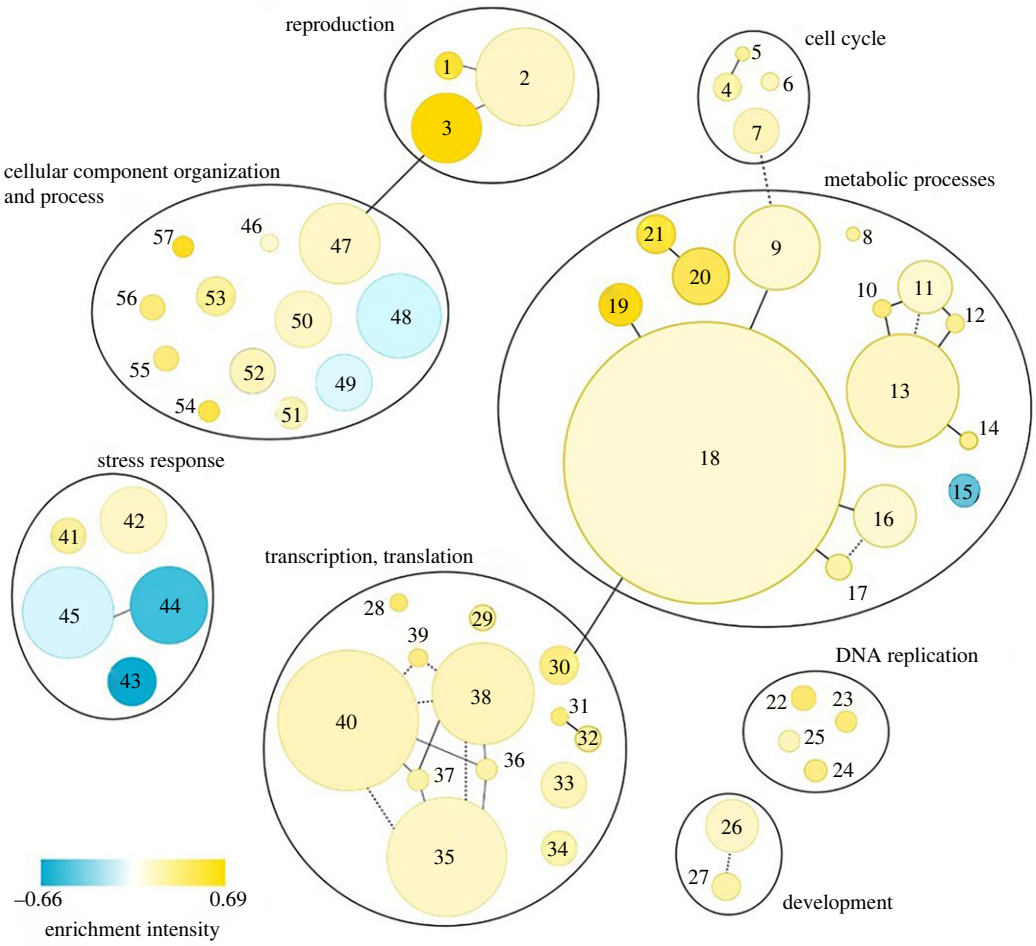

Figure 4. Biological processes dysregulated in transmissible neoplasia (p-value < 0.05): **1.**Meiosis I; **2.**Cellular process involved in reproduction; **3.**Sperm motility; **4.**Cytokenetic process; **5.**Contractil ring contraction; **6.**Mitotic spindle elongation; **7.**Regulation of cell cycle 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; **31.**Transcription initiation from RNA polymerase I promoter; **32.**DNA-templated transcription, initiation; **33.**Translational elongation; **34.**Spliceosomal snRNP assembly; **35.**mRNA processing; **36.**tRNA splicing, via endonucleolytic cleavage and ligation; **37.**tRNA wobble base modification; **38.**tRNA metabolic process; **39.**Deadenylation-dependent decapping of nuclear-transcribed mRNA; **40.**ncRNA metabolic process; **41.**Response to oxidative stress; **42.**Double-strand break repair; **43.**Cellular response to amino acid starvation; **44.**Response to bacterium; **45.**Defence response; **46.**Actin filament based process; **47.**Microtubule-based movement; **48.**Cell surface receptor signalling pathway; **49.**Anion transport; **50.**Golgi vesicle transport; **51.**Ribosomal small subunit assembly; **52.**Ribosomal large subunit assembly; **53.**rRNA-contain RNP complex; **54.**Protein secretion; **55.**Protein targeting to ER; **56.**Mitochondrial respiratory chain complex assembly; **57.**Protein insertion into ER membrane. The colour intensity is proportional to the enrichment intensity, expressed as the ratio between the number of transcripts that were significantly over- (yellow) or down-expressed (blue) compared with the total number of transcripts in the process. The disc area is proportional to the number of DETs linked to the process (from 2 to 191 transcripts). Lines connecting two functions represent shared transcripts (solid line means that all transcripts in the smaller family are part of the largest family; dotted line means that only some transcripts are shared).

2.3.1. Proliferation-related functions are altered

Some functions consistent with a high proliferative activity are altered in MtrBTN2 cells. The 'cell cycle' family (4.Cytokenetic process; 5.Contractil ring contraction; 6.Mitotic spindle elongation; 7.Regulation of cell cycle process, in figure 4) is enriched in MtrBTN2 cells, as well as 'DNA replication' (22.DNA replication initiation; 23.DNA strand elongation; 24.DNA replication, synthesis of RNA primer; 25.DNA-dependent DNA replication, in figure 4). More specifically, negative regulation of the cell cycle progression appeared to be reduced: *ASAH2*, which inhibits the ceramide mediation of the cell cycle arrest [60], was more highly expressed, whereas *PPP2R5B*, which inhibits cell cycle progression [61], was less expressed in MtrBTN2. Moreover, transcripts of genes promoting the cell cycle progression were more abundant in MtrBTN2 cells, in particular five genes that form the anaphase promoting complex (*ANAPC1, ANAPC4, ANAPC5, CDC26 and CDC23*) and *CCND1* encoding cyclin D1, which promotes progression through the G1-S phase of the cell cycle [62]. *TERT*, which is essential for the chromosome telomere replication in most eukaryotes and is particularly active in progenitor and cancer cells [63], was also more expressed in MtrBTN2 cells. By maintaining telomere length, telomerase removes a major barrier to cellular longevity, allowing unlimited proliferation driven by oncogenes [64]. In addition, in the 'development' family, we found a higher abundance of transcripts of genes involved in promoting cell growth, such as *TTK* [65], *METAP2* [66], *THBS1* (activator of TGF-β) [39], *MEAK7* [67] and

LAMTOR1 [68]. Transcripts of several genes that act as growth inhibitors were less abundant in MtrBTN2, as often seen in cancer contexts, such as *FAM107A* [69] and *LTBP3* [70], which keeps TGF-β in a latent state.

2.3.2. Metabolic pathways are modified

Many metabolic processes were shown to be altered (figure 4 and electronic supplementary material, table S6). Nucleotide metabolism in MtrBTN2 cells was characterized by a higher expression of more than 25 transcripts of genes involved in the de novo and salvage biosynthetic pathways of purines and pyrimidines and in the homeostasis of cellular nucleotides. The metabolism of protein and carbohydrate derivatives was also modified. The enrichment concerned transcripts of genes encoding ribosomal proteins, protein turnover and localization (especially deubiquitination), proteolysis, peptidyl amino acid modification, histone acetylation and lipid and protein glycosylation. In addition to *MEAK7* and *LAMTOR1*, transcripts of eight genes (*BMP2*, *KICS2*, *ITFG2*, *MIOS*, *NPRL2*, *RagA*, *RagD and SEC13*) involved in the mTORC1 amino acid sensing pathway (43.Cellular response to amino acid starvation, in figure 4) [71] were differentially expressed. Interestingly, in addition to the role of mTORC1 in the regulation of autophagy, mTORC1 plays a pivotal role as a master regulator of protein, lipid, and nucleic acid metabolism changes that have been reported in many different cancer cells [72–75]. Several transcripts of genes involved in organelle activities and related to protein metabolism (anabolism, catabolism, post-translational modifications, folding and secretion) were also more highly expressed in MtrBTN2.

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

2.3.3. DNA repair systems are dysregulated

MtrBTN2 cancer was found to express at higher levels a significant panel of transcripts of genes involved in the DNA homologous recombination (HR) (42.Double-strand break repair, in figure 4), such as *BABAM1*, *ZSWIM7*, *RAD51L*, *RAD54L*, *MRE11*, *MCM9* and *TONSL*. The DNA helicase MCM9 is also involved in the repair of inter-strand crosslinks [78], as is MUS81 [79], which is also more expressed in MtrBTN2 cells. This type of damage is a double-edged sword, as double-strand breaks can induce cell death if not efficiently repaired, but inefficient or inappropriate repair can lead to mutation, gene translocation and cancer [80]. Due to their high proliferation rate and increased metabolic activities, cancer cells are under a huge replication stress [81], which may emphasize the need for these long-lived cancer lineages to activate DNA repair to persist in time. Cells are able to repair double-strand breaks through two major pathways: non-homologous end joining and HR [82]. HR is critical for re-establishing replication forks at the sites of damage during S and G2 phases of the cell cycle. However, when combined with the depletion of cell cycle checkpoints and apoptosis, this mechanism can lead to genomic instability. As the ploidy of MtrBTN2 appears to be highly abnormal (varying from 8N to 18N [19]), further investigations will be required to determine whether the dysregulation of these DNA repair pathways favours recombination between homologous chromosomes and the accumulation of chromosomal abnormalities, or whether these pathways are activated in order to limit DNA damage. Interestingly, upregulation of double-stranded DNA break repair via the HR pathway has also been found in DFTD, in addition to repair via non-homologous end joining [83].

2.3.4. Interaction with extracellular matrix is affected

The expression levels of transcripts of seven genes encoding integrin subunits (48.Cell surface receptor signalling pathway, in figure 4), a major class of transmembrane glycoproteins that mediate cell–matrix and cell–cell adhesion, were altered (electronic supplementary material, table S6). As integrins are mostly less expressed in MtrBTN2 cells, this suggests that the ability of these cells to interact with ECM components is profoundly altered, which is often involved at multiple stages of the cancer processes [84].

2.3.5. Cell fate determination pathways are modified and soma-to-germline transition may have occurred

Major cell fate pathways appear to be modified (figure 4 and electronic supplementary material, table S6). The expression of transcripts of six genes involved in the Notch and Wnt pathways was altered. These changes may drive dedifferentiation processes often observed in aggressive cancer cells or cancer stem cells [85]. We report the results of the focused analysis performed on these two signalling pathways in §2.5. In addition, three biological processes clustered under GO terms related to 'reproduction' were enriched (1.Meiosis I; 2.Cellular process involved in reproduction; 3.Sperm motility, in figure 4). Nineteen transcripts more expressed in MtrBTN2 cells gathered under the 'sperm motility' process and encode dynein assembly factors, dynein chains and cilia and flagella-associated proteins. Moreover, transcripts of genes involved in meiosis such as *MSH5* [86], *SYCP3* [87] and *TEX11* [88] were more abundant in MtrBTN2. These unexpected results may reflect a soma-to-germline transition during the oncogenic process, which has been reported in a variety of cancers, and support the hypothesis of profound changes in the differentiation state of MtrBTN2 cells [89,90]. Interestingly, when abnormally produced in mitotically dividing cells, the *SYCP3* can impair recombination and drive ploidy changes that affect chromosome segregation in cancer cells, which may be one of the mechanisms that led to the hyperploidy of MtrBTN2 cells [90].

2.3.6. Immune response is altered

One of the most striking differences between MtrBTN2 cells and haemocytes is that biological processes belonging to innate immunity (44.Response to bacterium; 45.Defence

response, in figure 4) are significantly underrepresented in MtrBTN2 cells. Transcripts of genes encoding important mussel antimicrobial peptides and bactericidal proteins such as *Mytilin B*, *Myticin A*, *Myticin B*, *MGD1* and *BPI* [91–93] were significantly less abundant in MtrBTN2 cells. A similar pattern was also observed for genes encoding several pattern recognition receptors, such as *TLR2* and *TLR4*, which belong to the toll-like-receptor family and are essential for innate immunity against pathogens [94], and the lectin-related pattern recognition molecules *FCN2*, *FIBCD1*, *CLEC4E* and *CLEC4M*, which are involved in the recognition of pathogens and apoptotic or necrotic cells [95–97]. This underrepresentation of biological functions related to host defences suggests that MtrBTN2 cells are immunologically incompetent. Although MtrBTN2 cells share the same compartments in the mussel tissues as haemocytes, and a study of *Cerastoderma* BTNs suggests that they are derived from haemocytes [32], the tissue of origin of MtrBTN2 has not yet been confirmed. Indeed, this immunological incompetence could be due to a loss of immunological functions or be the result of a different tissue of origin. During MtrBTN2 disease progression, we also observe a drastic decrease in the number of host haemocytes, which are progressively replaced by circulating MtrBTN2 cells, which can represent over 95% of the circulating cells in cancerous mussels [19,22]. It remains to be determined whether MtrBTN2 cells outcompete normal haemocytes in the haemolymph or whether they interfere with haematopoiesis, as seen in some human leukaemias [98]. Still, the combination of fewer circulating haemocytes and the immune incompetence of MtrBTN2 cells could have dramatic consequences for the host health, leading to lethal opportunistic systemic infections [99].

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

Protein–protein interaction (PPI) network of the DETs was constructed to evaluate their connectivity and to identify hub genes that may play a critical role in the MtrBTN2 phenotype. The complete PPI network is shown in figure 5. The top 20 genes in the connectivity ranking were found in functions already identified as enriched with GO_MWU, such as translation (most involved in ribosome biogenesis) and DNA replication. Outputs from NetworkAnalyzer v4.4.8 [100] are reported in electronic supplementary material, table S7. Interestingly, this analysis highlighted three genes as major hubs in the PPI network that did not stand out in the GO_MWU analysis, namely *FN1*, *CDC42*, and *CASP3*. Fibronectin 1, encoded by *FN1*, is a major component of the ECM and plays an important role in cell adhesion, migration, growth and differentiation [101]. The abundance of *FN1* transcripts was significantly higher in MtrBTN2 cells (log2FC > 5). CDC42, a member of the Rho GTPase family, plays an important role in cell–cell and cell–matrix adhesion, actin cytoskeleton architecture, cell migration, and cell cycle regulation [102]; corresponding transcripts were significantly less abundant in MtrBTN2 cells. As Fibronectin 1 and CDC42 are involved in complementary functions and signalling pathways (such as cell–ECM interactions and cell migration), both hub genes were found to be interconnected in the PPI network. In addition, a large number of integrin subunits were found to be dysregulated. Taken together, these results strongly suggest that cell–ECM interactions and cell migration functions are profoundly modified in MtrBTN2 cells, consistent with their observed rounded morphology, poor adhesion properties and f-actin modifications [19,21]. The master effector of apoptosis, Caspase 3, encoded by *CASP3*, is also less expressed in MtrBTN2 cells. It plays a central role in the execution phase of cell apoptosis [103], which is often inhibited in cancer cells, allowing them to escape cell death, despite the accumulation of DNA damage and cell cycle dysfunction.

2.5. Most key oncogenic signalling pathways are altered in MtrBTN2

Based on the Cancer Genome Atlas project, Sanchez-Vega *et al*. [27] highlighted that 10 signalling pathways are very frequently altered in most human cancers. We found that six of these 10 oncogenic signalling pathways may be altered in MtrBTN2 cancer. These pathways are involved in cell proliferation and differentiation.

DETs were found for key genes related to the 'Hippo', 'Notch', 'Wnt', 'Myc', 'PI3K' and 'Cell Cycle' signalling pathways (table 3).

The Hippo signalling pathway plays a role in inhibiting cell proliferation, controlling cell fate and promoting apoptosis through the phosphorylation of YAP/TAZ transcription coactivators, their cytoplasmic retention and degradation. *SAV1* is a key activator of the Hippo pathway involved in LATS1/2 phosphorylation, a necessary step for subsequent YAP/TAZ phosphorylation [104]. As *SAV1* is less expressed in MtrBTN2 cells, the Hippo pathway may be downregulated or inactivated.

The PI3K pathway also showed dysregulation. It is involved in cell metabolism, growth, proliferation, cell–ECM interactions and survival. We observed a lower expression of two genes, *PTEN* and *PIK3R1*, which act as PIK3 antagonists [105]. When the pathway is activated, AKT1 is phosphorylated and inhibits the activity of TSC1/2. AKT-mediated phosphorylation of TSC1/2 removes its inhibition of RHEB activity, leading to activation of the mTORC1 complex. Our enrichment analysis (§2.3) has also highlighted the activation of mTORC1 by the alternative amino acid sensing pathway. In addition, AKT1 plays an indirect role in the Wnt and Myc pathways by negatively regulating GSK3β through its phosphorylation [106]. Thus, the inhibitory effect of unphosphorylated GSK3β on MYC and its contribution in the destruction complex of β-cathenin in the Wnt pathway may be repressed in MtrBTN2.

The gene encoding MAD3, an antagonist of MYC for MAX binding [107], was less expressed in MtrBTN2 cells, suggesting that the MYC–MAX complex is favoured and the pathway is activated.

Among SFRPs, which act as Wnt pathway inhibitors by binding extracellular WNT ligands, *SFRP1* was less expressed and *SRFP3* more. Moreover, *APC*, which encodes a constitutive protein of the destruction complex, was more expressed. APC is a multifunctional protein and is involved, for instance, in the normal compaction of mitotic chromatin [108]. We looked for differentially expressed genes among the downstream genes of the Wnt pathway, which transcription is regulated by unphosphorylated CTNNB1 mediation. We found that *CCND1*, *JAG1* and *FN1* were more expressed

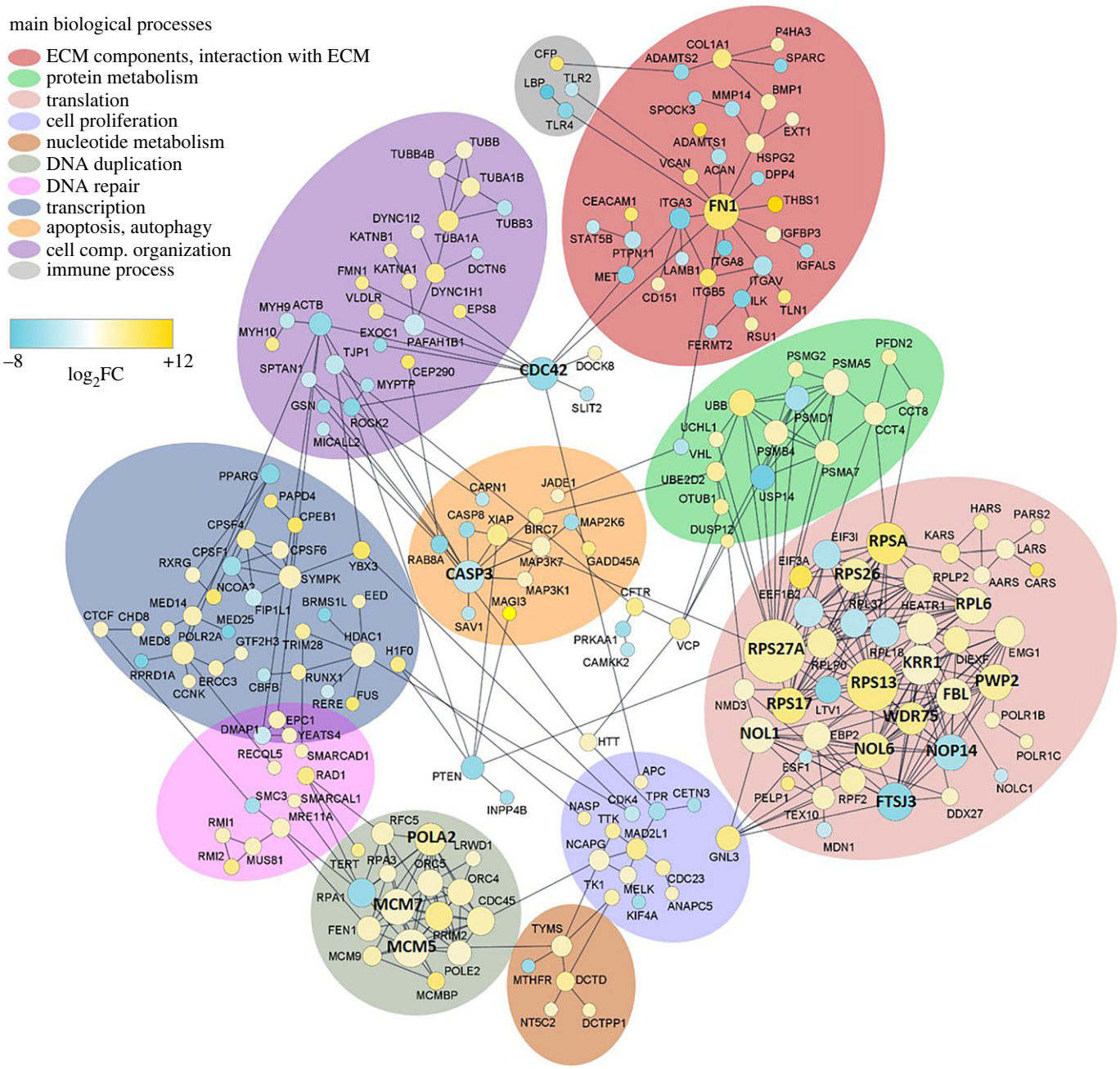

Figure 5. Protein–protein interaction (PPI) network. The size of the disc is proportional to the number of connections. The top 20 hub proteins are shown in bold. Up-regulated genes are shown in yellow; down-regulated genes are shown in blue. The colour intensity is proportional to the log2FC value.

suggesting that the Wnt pathway was activated [109]. CCND1 promotes cell proliferation and JAG1 is a transmembrane ligand that activates the Notch pathway. However, FBXW7, which inhibits the NOTCH cleavage and its subsequent activation, was more expressed in MtrBTN2 cells. In addition, the *HES1* expression, which is promoted by activated NOTCH, was lower in MtrBTN2 cells. HES1 negatively regulates the expression of downstream target genes such as tissue-specific 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 observed in all three MtrBTN2 samples (CAN1, 2, 3) but were completely absent in the 6 MtrBTN2-free samples (TRO1–3, EDU1–3). MtrBTN2-specific SNVs were identified in 8 key genes (table 4; for alignments, see electronic supplementary material, files S1 and S2) and were confirmed to be present in the additional cancerous individuals (CAN4–6).

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 molluscs [111] we found five MtrBTN2-specific SNVs. The variants T177C and C816T are synonymous substitutions previously described by Vassilenko *et al*. [112] in *M. trossulus* affected by haemic

Table 3. Genes differentially expressed in MtrBTN2 among the most frequently altered genes in the oncogenic signalling pathways. We report their action on the corresponding pathway.

pathway	main functions	gene	action on pathway	MtrBTN2/haemocytes expression	log2FC
Hippo	anti-proliferative	SAV1	activation	lower	−3.65
	anti-stem cell self renewal				
	pro-apoptotic				
Notch	cell proliferation	JAG1 (2 genes)	activation	higher	3.75 and 4.32
	cell fate	FBXW7	repression	higher	1.24
Wnt	cell proliferation	SFRP1	repression	lower	−3.54
	stemness maintenance	SFRP3	repression	higher	7.38
		APC	repression	higher	2.63
		CTNNB1	effector	higher	1.09
Myc	cell proliferation	MAD3	repression	lower	−2.19
	apoptosis				
PI3K	cell proliferation	PTEN	repression	lower	−4.51
	cell survival	PIK3R1	repression	lower	−1.68
Cell Cycle		CCND1	activation	higher	1.89

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.

			substitution type			
gene	transcript length	pathway	synonymous	missense	total number	missense substitution in a domain
PIK3CA	3144 bp	PI3K	33	0	33	0
PIK3CB	3159 bp	PI3K	3	0	3	0
PTEN	1395 bp	PI3K	2	0	2	0
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	0	1	0

neoplasia. The variant A1264T is a missense substitution, newly described here, but is not located in a domain. *Mytilus trossulus p53* is characterized by a deletion of 6 nucleotides if compared to *M. edulis p53* [113]. We found this deletion in 100% of reads coming from MtrBTN2 samples. The frequencies of these MtrBTN2-specific SNVs are reported in electronic supplementary material, table S8, and in most cases, heterozygosity was observed, with maintenance of expression of the *M. trossulus* alleles whereas host-specific (*M. edulis*) alleles were predominantly absent (electronic supplementary material, files S1). Indeed, the real impact of these variants on the cellular functions is difficult to assess and this MtrBTN2-specific variant set could include both somatic variants and ancestral germline polymorphisms.

3. Discussion

In their survey of bilaterian animals, Aktipis *et al*. [114] have revealed that all or almost all bilaterians are susceptible to cancer. Molluscs (Lophotrochozoa) are no exception [20,22]. Although a high degree of genotypic and phenotypic diversity has been described among mammalian cancers [115], molecular studies, mainly in humans and mice, have identified common drivers of the cancer process [26,27]. Still, oncogenic processes in invertebrate phyla have been poorly studied, especially at the molecular level. Even if expression levels are not the only factor determining the effect of genes on phenotypes, our present transcriptomic data suggest that the most common alterations associated with oncogenesis found in mammals are present in MtrBTN2 cells. Six of the major oncogenic signalling pathways are likely to be altered. One of the most fundamental traits of cancer cells is their high proliferation rate. We have previously found a mean doubling time of approximately 3 days for MtrBTN2 [24]. The present study provides information on the molecular bases of the high proliferation of MtrBTN2. The overexpression of these proliferation-related pathways is probably even more evident because we have used specialized cells (haemocytes) as a reference. At present, the issue of

haematopoiesis in bivalves is poorly understood; however, some circulating haemocytes are capable of dividing and expressing genes involved in cell proliferation, particularly during the immune response [116]. The proliferative activity of MtrBTN2 appears to be supported by metabolic rewiring, as is often seen in cancer cells [77]. Indeed, we highlight a possible central role for the Pi3K-AKT-mTORC1 pathway, which is a master regulator in coupling cell growth and amino acid, lipid, and nucleotide metabolisms that are frequently modified in many mammalian cancers of diverse origin [117–119]. Our data also suggest inhibition of apoptosis. Dysregulation of receptors and pathways involved in cell–ECM interactions [120] was also highlighted here, which is consistent with our knowledge of MtrBTN2 biology, pathogenesis (intra-host disease progression) and its inter-host transmission. Indeed, MtrBTN2 cells are non-adherent circulating cells that infiltrate tissues and breach physiological barriers such as epithelium during the transmission process. MtrBTN2 cells are also characterized by a dysregulated differentiation state often seen in aggressive cancers [85], as reflected by the aberrant activation of some meiosis-related genes suggesting a soma-to-germline transition. Such reactivation of meiosis-related gene expression has previously been observed in a variety of non-germ cell cancers in humans [89]. Finally, our transcriptomic data suggest a dampening of host defences in cancerous mussels, which is likely to facilitate disease progression and pathological outcome. Indeed, evasion from immune destruction by silencing components of the immune system is another hallmark of human cancers [26].

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

4. Material and methods

4.1. Mussel samplings and MtrBTN2 diagnostic

We collected mussels in the English Channel and the Barents Sea in late 2019. Two hundred *Mytilus edulis* were sampled in a farm located in Agon-Coutainville (49°0′44.784″ N, 1°35′55.643″ W, Normandy, France) and immediately screened for cancer at the LABEO laboratory (Caen, France). The presence of circulating cancer cells and the stage of the disease were initially diagnosed by cytological examination of haemolymph samples [19]. We found 14 positive individuals among which three were at an advanced stage of the disease (greater than 95% of circulating cells were cancer cells). After anaesthesia [121], we drew a maximum volume of haemolymph from the adductor muscle of these three cancerous mussels, as well as from three mussels diagnosed as MtrBTN2-free. The haemolymph was placed individually in RNase-free microtubes kept on ice and centrifuged at 800×*g* for 10 min at 4°C. The pelleted cells were immediately resuspended in TRIzol (Invitrogen) and stored at −80°C until RNA extraction. As non-transmissible circulating cancers also exist in mussels [20], we confirmed the MtrBTN2 diagnosis by two qPCRs, one specific for *M. trossulus* (and MtrBTN) and targeting the nuclear marker EF1*α* [24] and one specific for the MtrBTN2 lineage targeting mtCR [16]. Briefly, a piece of mantle and gills was fixed in 96% ethanol and used for DNA extraction using the Nucleomag 96 Tissue Kit (MachereyNagel). We carried out both qPCRs on cancerous and cancer-free *M. edulis* using the sensiFAST™-SyBR No-ROX Kit (Bioline) and the LightCycler 480 Real-Time PCR (Roche Life Science) system. A transversal section (including gills, digestive gland, mantle and gonad) of each individual was fixed in Davidson for 48 h and embedded in paraffin (RHEM facility). Sections were cut at a thickness of 3 µm and stained with haematoxylin and eosin to confirm the health status of each mussel and to exclude the presence of other pathologies.

Three *M. trossulus* were collected in the wild in Mishukovo (69°02′39;34.3″ N, 33°01′39;45.9″ E, Kola Bay, Russia, Barents Sea). The haemolymph was drawn as for *M. edulis* individuals, the cells were pelleted and resuspended in 100 µl of RNA*later* (Invitrogen) and sent to our laboratory where they were frozen at −80°C until RNA extraction. We also received a piece of mantle and gills from each individual fixed in 96% ethanol for genetic screening and to exclude the presence of MtrBTN, and tissue sections in 70% ethanol after 48h of fixation in Davidson's solution. *M. trossulus* individuals were subjected to both qPCRs and histological examination.

4.2. Total RNA extraction, library preparation, and sequencing

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

Polyadenylated RNA-seq library construction and sequencing using Illumina technology were performed by the GENEWIZ Company (Germany). Four hundred nanograms of total RNA at a concentration of 10 ng μl^{-1}, and OD260/280 comprised between 1.85 and 2.21 were used for library preparation. The NEBnext Ultra II Directional RNA kit was used for the cDNA library preparation and 9 cycles of enrichment PCR were run. Sequencing was performed

on Illumina NovaSeq, with a 150 bp paired-end configuration, and a sequencing depth of 100 M raw paired-end reads per sample.

4.3. Additional samples

Sixteen additional mussel individuals were used to validate our results. Ten mussels (5 MtrBTN2-affected and 5 MtrBTN2-free) were collected in a distinct location, at Le Croisic (47°17′58.5″ N, 2°31′00.4″ W, Loire-Atlantique, France) in February 2023 and used to validate RNAseq-based quantification and differential expression between MtrBTN2 cells and normal haemocytes by RT-qPCRs (see §4.5). Total RNA extraction was performed as previously described (see §4.2). RNAseq reads from 6 mussels (3 MtrBTN2-affected and 3 MtrBTN2-free) collected in Agon-Coutainville (49°0′44.784″ N, 1°35′55.643″ W, Normandy, France) in February 2021 for a separate study were used to confirm the differential transcriptomic profile of MtrBTN2 (see §4.5) and to validate the MtrBTN2-specific SNPs (see §4.6). Whole flesh from these 6 individuals was individually frozen in liquid nitrogen and powdered by cryogenic grinding (Mixer Mill MM 400, Retsch). For RNA extraction, mussel powder (10 mg) was homogenized in 500 µl of TRIzol by vortexing for 1 h at 4°C. Prior to extraction, performed as described above, insoluble material was removed by centrifugation at 12 000×*g* for 10 min at 4°C.

4.4. De novo transcriptome assembly and functional annotation

Raw reads were processed using RCorrector v1.04 (https://anaconda.org/bioconda/rcorrector/files?version=1.0.4) with default settings and -rf configuration to correct sequencing errors [122]. Uncorrectable reads were then removed using the FilterUncorrectabledPEfastq tool (https://github.com/harvardinformatics/TranscriptomeAssemblyTools/blob/master/FilterUncorrectabledPEfastq.py). The output reads were further processed for adapter removal and trimming with TrimGalore! v0.6.4 (https://github.com/FelixKrueger/TrimGalore) with default parameters and -q 28, –length 100. Ribosomal RNAs potentially still present after polyA capture were removed by alignment against the SILVA Ribosomal database with Bowtie2 v2.4.1 [123]. Read quality was assessed before and after read trimming using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

In France, MtrBTN2 infects *M. edulis* hosts, but the lineage originated in a *M. trossulus* founder host. These two species are closely related, hybridize when they come into contact, either naturally or via human-induced introductions [124–126], and introgression between them is observed. However, *M. trossulus* is not present in France and individuals of this species have been sampled in a different environment (Barents Sea versus English Channel for MtrBTN2). Therefore, to allow a differential gene expression analysis of the cancer cells, all retained reads (from MtrBTN2 cells, *M. edulis* haemocytes and *M. trossulus* haemocytes) were assembled into a pantranscriptome using Trinity v2.8.5 [127] with the default options. TransDecoder v3.0.1 (https://github.com/TransDecoder/TransDecoder/wiki) was run on these contigs to identify CDSs with a minimum length of 90 amino acids. Finally, CDS sequence redundancy was reduced using CD-HIT-EST v4.8.1 (https://github.com/weizhongli/cdhit/wiki) with the following options: -n 6, -c 0.86, -G 0, -aL 1, -aS 1. We assessed the quality of the assembly using several tests. The assembly was first analysed using TrinityStats, and the final pantranscriptome completeness was estimated using BUSCO v5.1.1 [128] against the conserved single-copy metazoan gene database ($n = 978$). Finally, filtered reads were mapped back to the filtered pantranscriptome to assess the individual mapping rate with Bowtie2 v2.4.1 [123]. To conclude that our pantranscriptome strategy was reliable, we performed a redundancy analysis (RDA) on the normalized read counts to analyse the impact of explanatory variables ('cell type': haemocyte/MtrBTN2 cells, 'species': trossulus/edulis, 'environment': Barents Sea/English Channel) on response variables (gene expression), followed by an ANOVA-like permutation test (nperm = 999, model = 'full') (https://cran.r-project.org/web/packages/vegan/vegan.pdf) (electronic supplementary material, figure S7). Both 'cell type' and 'environment' were retained as significant explanatory variables ($p < 0.05$). This confirms that the inclusion of healthy *M. edulis* mussels in the analysis is necessary to subtract the environment effect (English Channel versus Barents Sea).

For functional annotation, the transcripts were searched against the Uniprot (Swissprot and TrEMBL) protein reference database [129] using PLASTX v2.3.3 algorithm with an e-value cutoff of 1.0×10^{-3} [130]. Domain prediction against the InterPro database [131] was carried out with InterProScan v5.48–83.0. Both results were combined and we used the OmicsBox program v2.0 [34,35] to assign GO terms to the annotated sequences with an e-value hit filter of 1×10^{-3}, an annotation cutoff of 55, a HSP-Hit coverage of 60%, and an evidence code of 0.8. We searched for additional correspondences using EggNog v5.0 [132] by an orthology analysis.

4.5. Quantification of transcript expression

To investigate differential expression between MtrBTN2 cells and *M. edulis*–*M. trossulus* haemocytes, reads were first aligned to the pantranscriptome using SALMON v1.6 [133]. DETs were identified using DESeq2 v1.34 [36] with a FDR < 0.05. Count normalization was performed using the RLE method implemented in the DESeq2 package [134]. To confirm the identified DETs, we ran the QuantEval pipeline (https://github.com/dn070017/QuantEval, 37) with the 'contig mode' and DESeq2 on the quantification outputs. We compared the quantification results by Spearman's correlation analysis and we retained the shared DETs for the subsequent analyses.

To validate the quantifications and differential expression profiles obtained from our RNA-seq data, we carried out two alternative analyses on additional samples (see §4.3). First, we performed RT-qPCRs targeting 17 genes (*ADAMTS1*, *ALOX5*, *APC*, *BMP2*, *CCND1*, *CTNNB1*, *CYP11A1*, *FBXW7*, *JAG1*, *MAD3*, *PIK3R1*, *PTEN*, *SAV1*, *SELP*, *SFRP1*, *SFRP3* and *THBS1*) selected for their contrasting expression levels according to the DESeq2 analysis and their oncogenic relevance. Reverse transcription into cDNA was performed using M-MLV Reverse Transcriptase (MMLV-RT, Invitrogen) following the manufacturer's instructions. The total RT-qPCR reaction volume was 1.5 µl and consisted of 1 µl of LightCycler 480 SYBR Green I Master Mix (Roche) containing 0.5 µM of PCR primer (Eurogenetec) and 0.5 µl of cDNA (1/10 dilution). The amplification efficiency of each primer pair (electronic supplementary material, table S7) was validated

by serial dilution of a pool of all cDNAs. Relative expression was calculated as the threshold cycle (Ct) values of selected genes minus the mean of the measured threshold cycle (Ct) values of three constitutively expressed genes (*GAPDH*, *RPS28*, and *UBE2D3*) (electronic supplementary material, table S8). We compared the log2FC values obtained from the two quantification methods. Log2FCs for RT-qPCRs were calculated as follows:

$$FC = \text{efficiency of the qPCR}^{\text{mean of Ct of the 5 MtrBTN2 affected individuals} - \text{mean of Ct of the 5 MtrBTN2 free individuals}}.$$

We also performed Spearman's correlation analysis between normalized expression obtained from RNA-seq data (RLE values) and normalized Ct obtained from RT-qPCR.

Second, we performed DESeq2 analysis on complementary RNA-seq data obtained from the whole flesh of 6 additional samples from a separate study (see §4.3) to confirm transcripts specific to the malignant state.

4.6. Biological interpretation of gene expression profiles

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

4.6.1. Enrichment analysis

We performed a GO term enrichment analysis focusing on biological processes with the GO_MWU tool (https://github.com/z0on/GO_MWU) using adaptive clustering and a rank-based statistical test (Mann–Whitney U-test). We used the following parameters for adaptive clustering: largest = 0.5; smallest = 10; clusterCutHeight = 0.5. We considered both the level of expression and the significance of the differential expression: we assigned the log2 fold change value to transcripts that were significantly differentially expressed (adjusted $p < 0.05$), while assigning zero to the others. We considered as enriched a biological process with a FDR < 1%. To display the results synthetically, we used the Enrichment Map v3.3.3 tool [135] in Cytoscape v3.9.1 [136]. The intensity of the enrichment was evidenced in the network and was calculated as follows: (i) for the processes enriched with over-expressed transcripts, 'number of transcripts significantly over-expressed in the process/total number of transcripts in the process'; (ii) for the processes enriched with under-expressed transcripts, '−1 × (number of transcripts significantly downregulated in the process/total number of transcripts in the process)'.

4.6.2. Hub and top expressed gene identification

The top 20 genes in terms of connectivity ranking in the PPI network were selected as hub genes. We used the Search Tool for the Retrieval of Interacting Genes (STRING), a biological database designed to predict PPI networks [137]. The results were visualized in Cytoscape v3.9.1 using NetworkAnalyzer v4.4.8 visualization software [100], which can construct comprehensive models of biological interactions. Isolated and partially connected nodes were not included.

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

4.6.3. Targeted analyses

We performed a targeted analysis focusing on genes and pathways that have been identified by Sanchez-Vega *et al.* [27] as being altered at high frequency across many different human cancer types. We searched for the presence of a differential expression of these genes in MtrBTN2 cells or of MtrBTN2-specific alleles (i.e. present only in MtrBTN2-affected mussels, and present in all MtrBTN2-affected mussels). We also identified *M. edulis* and *M. trossulus*-specific alleles. Allele visualization and calculation of variant frequencies were carried out using the IGV v2.12.3 [138] and iVar [139] tools on BAM files obtained after read alignment to the transcript sequences with Bowtie2 v2.4.1 [123]. The minimum quality score threshold to count a base and the minimum frequency threshold were set to 20 and 0.02, respectively, with a minimum base coverage of 50 reads. MtrBTN2-specific alleles were searched in CAN1–3 individuals and then confirmed in additional CAN4–6 individuals (see §4.3).

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Raw read sequences were deposited with links to BioProject accession number PRJNA749800 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

Additional data are provided in electronic supplementary material [140].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. E.A.V.B.: conceptualization, formal analysis, investigation, methodology, resources, writing—original draft; M.H.: formal analysis, investigation, resources, writing—review and editing; E.V.: conceptualization, methodology, writing—review and editing; J.V.-D.: conceptualization, methodology, writing—review and editing; G.M.: conceptualization, writing—review and editing; F.T.: funding acquisition, writing—review and editing; N.B.: conceptualization, funding acquisition, resources, writing—review and editing; D.D.-G.: conceptualization, funding acquisition, writing—review and editing; G.M.C.: conceptualization, funding acquisition, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. The authors declare that they have no competing interests.

Funding. This work was supported by the French National Agency for Research, TRANCAN project (grant no. ANR-18-CE35-0009) and the Montpellier Université d'Excellence (MUSE), BLUECANCER project. F.T. is supported by the MAVA Foundation.

Acknowledgments. We acknowledge the Animal Health-Clinical Biology unit of the LABÉO laboratory (Caen, France), especially Maryline Houssin and Ludovic Petit, for their help during cytological screening and Professor Petr Strelkov from the Saint Petersburg State University for the sampling of *M. trossulus* individuals. We thank the Réseau d'Histologie Expérimentale de Montpellier, RHEM facility, supported by SIRIC Montpellier Cancer (grant INCa_Inserm_DGOS_12553), the European regional development foundation and the Occitanian region (FEDER-FSE 2014-2020 Languedoc Roussillon), for processing our animal tissues, histology technics and expertise. This study is set within the framework of the 'Laboratoires d'Excellence (LABEX)' Tulip (ANR- 10-LABX-41).

References

1. Ostrander EA, Davis BW, Ostrander GK. 2016 Transmissible tumors: breaking the cancer paradigm. *Trends Genet.* **32**, 1–15. (doi:10.1016/j.tig.2015.10.001)
2. Dujon AM *et al.* 2020 Transmissible cancers in an evolutionary perspective. *iScience* **23**, 101269. (doi:10.1016/j.isci.2020.101269)
3. Nowinsky M. 1876 Zurfrageueber die Impfung der krebsigengeschwuelste. *Zentralbl Med Wissensch.* **14**, 790–791.
4. Murgia C, Pritchard JK, Kim SY, Fassati A, Weiss RA. 2006 Clonal origin and evolution of a transmissible cancer. *Cell* **126**, 477–487. (doi:10.1016/j.cell.2006.05.051)
5. Baez-Ortega A *et al.* 2019 Somatic evolution and global expansion of an ancient transmissible cancer lineage. *Science* **365**, eaau9923. (doi:10.1126/science.aau9923)
6. Pye RJ *et al.* 2016 A second transmissible cancer in Tasmanian devils. *Proc. Natl Acad. Sci. USA* **113**, 374–379. (Doi:10.1073/pnas.1519691113)
7. Siddle HV *et al.* 2007 Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proc. Natl Acad. Sci. USA* **104**, 16 221–16 226. (doi:10.1073/pnas.0704580104)
8. Hamede RK, Bashford J, McCallum H, Jones M. 2009 Contact networks in a wild Tasmanian devil (*Sarcophilus harrisii*) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecol. Lett.* **12**, 1147–1157. (doi:10.1111/j.1461-0248.2009.01370.x)
9. Hamede RK, McCallum H, Jones M. 2013 Biting injuries and transmission of Tasmanian devil facial tumour disease. *J. Anim. Ecol.* **82**, 182–190. (doi:10.1111/j.1365-2656.2012.02025.x)
10. Hawkins CE *et al.* 2006 Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biol. Conserv.* **131**, 307–324. (doi:10.1016/j.biocon.2006.04.010)
11. Lachish S, Jones M, McCallum H. 2007 The impact of disease on the survival and population growth rate of the Tasmanian devil. *J. Anim. Ecol.* **76**, 926–936. (doi:10.1111/j.1365-2656.2007.01272.x)
12. McCallum H, Jones M, Hawkins C, Hamede R, Lachish S, Sinn DL, Beeton N, Lazenby B. 2009 Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. *Ecology* **90**, 3379–3392. (doi:10.1890/08-1763.1)
13. Lazenby BT *et al.* 2018 Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils. *J. Appl. Ecol.* **55**, 1368–1379. (doi:10.1111/1365-2664.13088)
14. Metzger MJ, Reinisch C, Sherry J, Goff SP. 2015 Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell* **161**, 255–263. (doi:10.1016/j.cell.2015.02.042)
15. Metzger MJ, Villalba A, Carballal MJ, Iglesias D, Sherry J, Reinisch C, Muttray AF, Baldwin SA, Goff SP. 2016 Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature* **534**, 705–709. (doi:10.1038/nature18599)
16. Yonemitsu MA *et al.* 2019 A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe. *Elife* **8**, e47788. (doi:10.7554/eLife.47788)
17. Garcia-Souto D *et al.* 2022 Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the seas of southern Europe. *Elife* **11**, e66946. (doi:10.7554/eLife.66946)
18. Riquet F, Simon A, Bierne N. 2017 Weird genotypes? Don't discard them, transmissible cancer could be an explanation. *Evol. Appl.* **10**, 140–145. (doi:10.1111/eva.12439)
19. Burioli EAV, Trancart S, Simon A, Bernard I, Charles M, Oden E, Bierne N, Houssin M. 2019 Implementation of various approaches to study the prevalence, incidence and progression of disseminated neoplasia in mussel stocks. *J. Invertebr. Pathol.* **168**, 107271. (doi:10.1016/j.jip.2019.107271)
20. Hammel M *et al.* 2022 Prevalence and polymorphism of a mussel transmissible cancer in Europe. *Mol. Ecol.* **31**, 736–751. (doi:10.1111/mec.16052)
21. Skazina M, Odintsova N, Maiorova M, Ivanova A, Väinölä R, Strelkov P. 2021 First description of a widespread *Mytilus trossulus*-derived bivalve transmissible cancer lineage in *M. trossulus* itself. *Sci. Rep.* **11**, 5809. (doi:10.1038/s41598-021-85098-5)
22. Carballal MJ, Barber BJ, Iglesias D, Villalba A. 2015 Neoplastic diseases of marine bivalves. *J. Invertebr. Pathol.* **131**, 83–106. (doi:10.1016/J.JIP.2015.06.004)
23. Benadelmouna A, Saunier A, Ledu C, Travers MA, Morga B. 2018 Genomic abnormalities affecting mussels (*Mytilus edulis-galloprovincialis*) in France are related to ongoing neoplastic processes, evidenced by dual flow cytometry and cell monolayer analyses. *J. Invertebr. Pathol.* **157**, 45–52. (doi:10.1016/j.jip.2018.08.003)
24. Burioli EAV, Hammel M, Bierne N, Thomas F, Houssin M, Destoumieux-Garzón D, Charrière GM. 2021 Traits of a mussel transmissible cancer are reminiscent of a parasitic life style. *Sci. Rep.* **11**, 24110. (doi:10.1038/s41598-021-03598-w)
25. Giersch RM *et al.* 2022 Survival and detection of bivalve transmissible neoplasia from the soft-shell clam *Mya arenaria* (MarBTN) in seawater. *Pathogens* **11**, 283. (doi:10.3390/pathogens11030283)
26. Hanahan D. 2022 Hallmarks of cancer: new dimensions. *Cancer Discov.* **12**, 31–46. (doi:10.1158/2159-8290.CD-21-1059)
27. Sanchez-Vega F *et al.* 2018 Oncogenic signaling pathways in the cancer genome atlas. *Cell* **173**, 321–337.e10. (doi:10.1016/j.cell.2018.03.035)
28. Sotiriou C, Piccart MJ. 2007 Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat. Rev. Cancer* **7**, 545–553. (doi:10.1038/nrc2173)
29. Wang Z, Gerstein M, Snyder M. 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63. (doi:10.1038/nrg2484)
30. Conesa A *et al.* 2016 A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (erratum: **17**, 181). (doi:10.1186/s13059-016-0881-8)
31. Martin JA, Wang Z. 2011 Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12**, 671–682. (doi:10.1038/nrg3068)
32. Bruzos AL *et al.* 2022 The evolution of two transmissible leukaemias colonizing the coasts of Europe. *BioRxiv.* 2022.08.06.503021. (doi:10.1101/2022.08.06.503021)
33. Payá-Milans M, Olmstead JW, Nunez G, Rinehart TA, Staton M. 2018 Comprehensive evaluation of RNA-seq analysis pipelines in diploid and polyploid species. *Gigascience* **7**, giy132. (doi:10.1093/gigascience/giy132)
34. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005 Blast2GO: a universal tool for annotation, visualisation and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676. (doi:10.1093/bioinformatics/bti610)
35. Gotz S *et al.* 2008 High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435. (doi:10.1093/nar/gkn176)
36. Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. (Doi:10.1186/s13059-014-0550-8)
37. Hsieh PH, Oyang YJ, Chen CY. 2019 Effect of de novo transcriptome assembly on transcript quantification. *Sci. Rep.* **9**, 8304. (doi:10.1038/s41598-019-44499-3)
38. Hu J, Zhang Z, Shen WJ, Azhar S. 2010 Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones. *Nutr. Metab. (Lond).* **7**, 47. (doi:10.1186/1743-7075-7-47)
39. Murphy-Ullrich JE, Poczatek M. 2000 Activation of latent TGF-β by thrombospondin-1: mechanisms and physiology. *Cytokine Growth Factor Rev.* **11**, 59–69. (doi:10.1016/S1359-6101(99)00029-5)
40. Letterio JJ, Roberts AB. 1998 Regulation of immune responses by TGF-β. *Annu. Rev. Immunol.* **16**, 137–161. (doi:10.1146/annurev.immunol.16.1.137)
41. Ashcroft GS. 1999 Bidirectional regulation of macrophage function by TGF-β. *Microbes Infect.* **1**, 1275–1282. (doi:10.1016/S1286-4579(99)00257-9)
42. Läubli H, Borsig L. 2010 Selectins promote tumor metastasis. *Semin. Cancer Biol.* **20**, 169–177. (doi:10.1016/j.semcancer.2010.04.005)
43. Fabricius HÅ, Starzonek S, Lange T. 2021 The role of platelet cell surface P-selectin for the direct

platelet-tumor cell contact during metastasis formation in human tumors. *Front. Oncol.* **11**, 642761. (doi:10.3389/fonc.2021.642761)

44. Anderson KM, Seed T, Vos M, Mulshine J, Meng J, Alrefai W, Ou D, Harris JE. 1998 5-Lipoxygenase inhibitors reduce PC-3 cell proliferation and initiate nonnecrotic cell death. *Prostate* **37**, 161–173. (doi:10.1002/(sici)1097-0045(19981101)37:3<161::aid-pros5>3.0.co;2-d)
45. Bishayee K, Khuda-Bukhsh AR. 2013 5-Lipoxygenase antagonist therapy: a new approach towards targeted cancer chemotherapy. *Acta Biochim. Biophys. Sin.* **45**, 709–719. (doi:10.1093/abbs/gmt064)
46. Lee JH, Lee GT, Woo SH, Ha YS, Kwon SJ, Kim WJ, Kim IY. 2013 BMP-6 in renal cell carcinoma promotes tumor proliferation through IL-10-dependent M2 polarization of tumor-associated macrophages. *Cancer Res.* **73**, 3604–3614. (doi:10.1158/0008-5472.CAN-12-4563)
47. Redondo-García S, Peris-Torres C, Caracuel-Peramos R, Rodríguez-Manzaneque JC. 2020 ADAMTS proteases and the tumor immune microenvironment: lessons from substrates and pathologies. *Matrix Biol. Plus* **9**, 100054. (doi:10.1016/j.mbplus.2020.100054)
48. Rucci N, Sanità P, Angelucci A. 2011 Roles of metalloproteases in metastatic niche. *Curr. Mol. Med.* **11**, 609–622. (doi:10.2174/156652411797536705)
49. Huang T, Sun L, Yuan X, Qiu H. 2017 Thrombospondin-1 is a multifaceted player in tumor progression. *Oncotarget* **8**, 84 546–84 558. (doi:10.18632/oncotarget.19165)
50. Bach DH, Park HJ, Lee SK. 2017 The dual role of bone morphogenetic proteins in cancer. *Mol. Ther. Oncolytics* **8**, 1–13. (doi:10.1016/j.omto.2017.10.002)
51. Yee KO, Connolly CM, Duquette M, Kazerounian S, Washington R, Lawler J. 2009 The effect of thrombospondin-1 on breast cancer metastasis. *Breast Cancer Res. Treat.* **114**, 85–96. (doi:10.1007/s10549-008-9992-6)
52. Nucera C *et al.* 2010 B-Raf(V600E) and thrombospondin-1 promote thyroid cancer progression. *Proc. Natl Acad. Sci. USA* **107**, 10 649–10 654. (doi:10.1073/pnas.1004934107)
53. Borsotti P *et al.* 2015 Thrombospondin-1 is part of a Slug-independent motility and metastatic program in cutaneous melanoma, in association with VEGFR-1 and FGF-2. *Pigment Cell Melanoma Res.* **28**, 73–81. (doi:10.1111/pcmr.12319)
54. Huang T, Wang L, Liu D, Li P, Xiong H, Zhuang L, Sun L, Yuan X, Qiu H. 2017 FGF7/FGFR2 signal promotes invasion and migration in human gastric cancer through upregulation of thrombospondin-1. *Int. J. Oncol.* **50**, 1501–1512. (doi:10.3892/ijo.2017.3927)
55. Iwamura T, Caffrey TC, Kitamura N, Yamanari H, Setoguchi T, Hollingsworth MA. 1997 P-selectin expression in a metastatic pancreatic tumor cell line (SUIT-2). *Cancer Res.* **57**, 1206–1212.
56. Ding XZ, Kuszynski CA, El-Metwally TH, Adrian TE. 1999 Lipoxygenase inhibition induced apoptosis, morphological changes, and carbonic anhydrase expression in human pancreatic cancer cells. *Biochem. Biophys. Res. Commun.* **266**, 392–399. (doi:10.1006/bbrc.1999.1824)
57. Gupta S, Srivastava M, Ahmad N, Sakamoto K, Bostwick DG, Mukhtar H. 2001 Lipoxygenase-5 is overexpressed in prostate adenocarcinoma. *Cancer* **91**, 737–743. (doi:10.1002/1097-0142(20010215)91:4<737::aid-cncr1059>3.0.co;2-f)
58. Ohd JF, Nielsen CK, Campbell J, Landberg G, Löfberg H, Sjölander A. 2003 Expression of the leukotriene D4 receptor CysLT1, COX-2, and other cell survival factors in colorectal adenocarcinomas. *Gastroenterology* **124**, 57–70. (doi:10.1053/gast.2003.50011)
59. Skovrlj B, Koehler SM, Anderson PA, Qureshi SA, Hecht AC, Iatridis JC, Cho SK. 2015 Association between BMP-2 and carcinogenicity. *Spine (Phila Pa 1976)* **40**, 1862–1871. (doi:10.1097/BRS.0000000000001126)
60. Wu BX, Zeidan YH, Hannun YA. 2009 Downregulation of neutral ceramidase by gemcitabine: implications for cell cycle regulation. *Biochim.. Biophys Acta* **1791**, 730–739. (doi:10.1016/j.bbalip.2009.03.012)
61. Loveday C *et al.* 2015 Mutations in the PP2A regulatory subunit B family genes PPP2R5B, PPP2R5C and PPP2R5D cause human overgrowth. *Hum. Mol. Genet.* **24**, 4775–4779. (doi:10.1093/hmg/ddv182)
62. Fu M, Wang C, Li Z, Sakamaki T, Pestell RG. 2004 Cyclin D1: normal and abnormal functions. *Endocrinology* **145**, 5439–5447. (doi:10.1210/en.2004-0959)
63. Pfeiffer V, Lingner J. 2013 Replication of telomeres and the regulation of telomerase. *Cold Spring Harb. Pers. Biol.* **5**, a010405. (doi:10.1101/cshperspect.a010405)
64. Bell RJ, Rube HT, Xavier-Magalhães A, Costa BM, Mancini A, Song JS, Costello JF. 2016 Understanding TERT promoter mutations: a common path to immortality. *Mol. Cancer Res.* **14**, 315–323. (doi:10.1158/1541-7786.MCR-16-0003)
65. Mills GB *et al.* 1992 Expression of TTK, a novel human protein kinase, is associated with cell proliferation. *J. Biol. Chem.* **267**, 16 000–16 006. (doi:10.1016/S0021-9258(19)49633-6)
66. Tucker L *et al.* 2008 Ectopic expression of methionine aminopeptidase-2 causes cell transformation and stimulates proliferation. *Oncogene* **27**, 3967–3976. (doi:10.1038/onc.2008.14)
67. Nguyen JT, Ray C, Fox AL, Mendonça DB, Kim JK, Krebsbach PH. 2018 Mammalian EAK-7 activates alternative mTOR signaling to regulate cell proliferation and migration. *Sci. Adv.* **4**, eaao5838. (doi:10.1126/sciadv.aao5838)
68. Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. 2010 Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* **141**, 290–303. (doi:10.1016/j.cell.2010.02.024)
69. Pastuszak-Lewandoska D *et al.* 2015 Decreased FAM107A expression in patients with non-small cell lung cancer. *Adv. Exp. Med. Biol.* **852**, 39–48. (doi:10.1007/5584_2014_109)
70. Saharinen J, Hyytiäinen M, Taipale J, Keski-Oja J. 1999 Latent transforming growth factor-beta binding proteins (LTBPs)-structural extracellular matrix proteins for targeting TGF-beta action. *Cytokine Growth Factor Rev.* **10**, 99–117. (doi:10.1016/s1359-6101(99)00010-6)
71. Shimobayashi M, Hall M. 2016 Multiple amino acid sensing inputs to mTORC1. *Cell Res.* **26**, 7–20. (doi:10.1038/cr.2015.146)
72. Cargnello M, Tcherkezian J, Roux PP. 2015 The expanding role of mTOR in cancer cell growth and proliferation. *Mutagenesis* **30**, 169–176. (doi:10.1093/mutage/geu045)
73. Ricoult SJ, Yecies JL, Ben-Sahra I, Manning BD. 2016 Oncogenic PI3K and K-Ras stimulate de novo lipid synthesis through mTORC1 and SREBP. *Oncogene* **35**, 1250–1260. (doi:10.1038/onc.2015.179)
74. Valvezan AJ *et al.* 2017 mTORC1 couples nucleotide synthesis to nucleotide demand resulting in a targetable metabolic vulnerability. *Cancer Cell* **32**, 624–638.e5. (doi:10.1016/j.ccell.2017.09.013)
75. Ben-Sahra I, Hoxhaj G, Ricoult SJH, Asara JM, Manning BD. 2016 mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. *Science* **351**, 728–733. (doi:10.1126/science.aad0489)
76. Martínez-Reyes I, Chandel NS. 2020 Mitochondrial TCA cycle metabolites control physiology and disease. *Nat. Commun.* **11**, 102. (doi:10.1038/s41467-019-13668-3)
77. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. 2008 The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab.* **7**, 11–20. (doi:10.1016/j.cmet.2007.10.002)
78. Nishimura K, Ishiai M, Horikawa K, Fukagawa T, Takata M, Takisawa H, Kanemaki MT. 2012 Mcm8 and Mcm9 form a complex that functions in homologous recombination repair induced by DNA interstrand crosslinks. *Mol. Cell* **47**, 511–522. (doi:10.1016/j.molcel.2012.05.047)
79. Hanada K, Budzowska M, Modesti M, Maas A, Wyman C, Essers J, Kanaar R. 2006 The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *EMBO J.* **25**, 4921–4949. (doi:10.1038/sj.emboj.7601344)
80. Chatterjee N, Walker GC. 2017 Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **58**, 235–263. (doi:10.1002/em.22087)
81. Di Micco R *et al.* 2006 Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642. (doi:10.1038/nature05327)
82. Haber JE. 2000 Partners and pathways repairing a double-strand break. *Trends Genet.* **16**, 259–264. (doi:10.1016/s0168-9525(00)02022-9)

83. Kozakiewicz CP *et al.* 2021 Spatial variation in gene expression of Tasmanian devil facial tumors despite minimal host transcriptomic response to infection. *BMC Genomics* **22**, 698. (doi:10.1186/s12864-021-07994-4)
84. Mizejewski GJ. 1999 Role of integrins in cancer: survey of expression patterns. *Proc. Soc. Exp. Biol. Med.* **222**, 124–138. (doi:10.1046/j.1525-1373.1999.d01-122.x)
85. Jilkine A, Gutenkunst RN. 2014 Effect of dedifferentiation on time to mutation acquisition in stem cell-driven cancers. *PLoS Comput. Biol.* **10**, e1003481. (doi:10.1371/journal.pcbi.1003481)
86. Bocker T *et al.* 1999 hMSH5: a human MutS homologue that forms a novel heterodimer with hMSH4 and is expressed during spermatogenesis. *Cancer Res.* **59**, 816–822.
87. Syrjänen JL, Pellegrini L, Davies OR. 2014 A molecular model for the role of SYCP3 in meiotic chromosome organisation. *Elife* **3**, e02963. (doi:10.7554/eLife.02963)
88. Adelman CA, Petrini JH. 2008 ZIP4H (TEX11) deficiency in the mouse impairs meiotic double strand break repair and the regulation of crossing over. *PLoS Genet.* **4**, e1000042. (doi:10.1371/journal.pgen.1000042)
89. McFarlane RJ, Wakeman JA. 2017 Meiosis-like functions in oncogenesis: a new view of cancer. *Cancer Res.* **77**, 5712–5716. (doi:10.1158/0008-5472.CAN-17-1535)
90. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. 2005 Cancer/testis antigens, gametogenesis and cancer. *Nat. Rev. Cancer* **5**, 615–625. (doi:10.1038/nrc1669)
91. Hubert F, Noel T, Roch P. 1996 A member of the arthropod defensing family from edible Mediterranean mussels (*Mytilus galloprovincialis*). *Eur. J. Biochem.* **240**, 302–306. (doi:10.1111/j.1432-1033.1996.0302h.x)
92. Mitta G, Hubert F, Noël T, Roch P. 1999 Myticin, a novel cysteine-rich antimicrobial peptide isolated from haemocytes and plasma of the mussel *Mytilus galloprovincialis*. *Eur. J. Biochem.* **265**, 71–78. (doi:10.1046/j.1432-1327.1999.00654.x)
93. Krasity BC, Troll JV, Weiss JP, McFall-Ngai MJ. 2011 LBP/BPI proteins and their relatives: conservation over evolution and roles in mutualism. *Biochem. Soc. Trans.* **39**, 1039–1044. (doi:10.1042/BST0391039)
94. Zhang L, Li L, Zhang G. 2011 A *Crassostrea gigas* Toll-like receptor and comparative analysis of TLR pathway in invertebrates. *Fish Shellfish Immunol.* **30**, 653–660. (doi:10.1016/j.fsi.2010.12.023)
95. Garlatti V *et al.* 2007 Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J.* **26**, 623–633. (doi:10.1038/sj.emboj.7601500)
96. Schlosser A, Thomsen T, Moeller JB, Nielsen O, Tornøe I, Mollenhauer J, Moestrup SK, Holmskov U. 2009 Characterization of FIBCD1 as an acetyl group-binding receptor that binds chitin. *J. Immunol.* **183**, 3800–3809. (doi:10.4049/jimmunol.0901526)
97. Dheilly NM *et al.* 2015 A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. *Dev. Comp. Immunol.* **48**, 234–243. (doi:10.1016/j.dci.2014.10.009)
98. Trino S, Laurenzana I, Lamorte D, Calice G, De Stradis A, Santodirocco M, Sgambato A, Caivano A, De Luca L. 2022 Acute myeloid leukemia cells functionally compromise hematopoietic stem/progenitor cells inhibiting normal hematopoiesis through the release of extracellular vesicles. *Front. Oncol.* **12**, 824562. (doi:10.3389/fonc.2022.824562)
99. Kent ML, Elston RA, Wilkinson MT, Drum AS. 1989 Impaired defense mechanisms in bay mussels, *Mytilus edulis*, with hemic neoplasia. *J. Invertebr. Pathol.* **53**, 378–386. (doi:10.1016/0022-2011(89)90103-1)
100. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. 2008 Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284. (doi:10.1093/bioinformatics/btm554)
101. Pankov R, Yamada KM. 2002 Fibronectin at a glance. *J. Cell Sci.* **115**, 3861–3863. (doi:10.1242/jcs.00059)
102. Cotteret S, Chernoff J. 2002 The evolutionary history of effectors downstream of Cdc42 and Rac. *Genome Biol.* **3**, REVIEWS0002. (doi:10.1186/gb-2002-3-2-reviews0002)
103. Porter AG, Jänicke RU. 1999 Emerging roles of caspase-3 in apoptosis. *Cell Death Differ.* **6**, 99–104. (doi:10.1038/sj.cdd.4400476)
104. Yu FX, Guan KL. 2013 The Hippo pathway: regulators and regulations. *Genes Dev.* **27**, 355–371. (doi:10.1101/gad.210773.112)
105. Carracedo A, Pandolfi P. 2008 The PTEN–PI3K pathway: of feedbacks and cross-talks. *Oncogene* **27**, 5527–5541. (doi:10.1038/onc.2008.247)
106. Cohen P, Frame S. 2001 The renaissance of GSK3. *Nat. Rev. Mol. Cell Biol.* **2**, 769–776. (doi:10.1038/35096075)
107. Amati B, Land H. 1994 Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr. Opin. Genet. Dev.* **4**, 102–108. (doi:10.1016/0959-437x(94)90098-1)
108. Dikovskaya D, Khoudoli G, Newton IP, Chadha GS, Klotz D, Visvanathan A, Lamond A, Swedlow JR, Näthke IS. 2012 The adenomatous polyposis coli protein contributes to normal compaction of mitotic chromatin. *PLoS ONE* **7**, e38102. (doi:10.1371/journal.pone.0038102)
109. Lecarpentier Y, Schussler O, Hébert JL, Vallée A. 2019 Multiple targets of the canonical WNT/β-catenin signaling in cancers. *Front. Oncol.* **9**, 1248. (doi:10.3389/fonc.2019.01248)
110. Iso T, Kedes L, Hamamori Y. 2003 HES and HERP families: multiple effectors of the Notch signaling pathway. *J. Cell Physiol.* **194**, 237–255. (doi:10.1002/jcp.10208)
111. Kelley ML, Winge P, Heaney JD, Stephens RE, Farell JH, Beneden RJV, Reinisch CL, Lesser MP, Walker CW. 2001 Expression of homologues for p53 and p73 in the softshell clam (*Mya arenaria*), a naturally-occuring model for human cancer. *Oncogene* **20**, 748–758. (doi:10.1038/sj.onc.1204144)
112. Vassilenko EI, Muttray AF, Schulte PM, Baldwin SA. 2010 Variations in p53-like cDNA sequence are correlated with mussel haemic neoplasia: a potential molecular-level tool for biomonitoring. *Mutat. Res.* **701**, 145–152. (doi:10.1016/j.mrgentox.2010.06.001)
113. Muttray AF, Cox RL, Reinisch CL, Baldwin SA. 2007 Identification of DeltaN isoform and polyadenylation site choice variants in molluscan p63/p73-like homologues. *Mar. Biotechnol. (NY)* **9**, 217–230. (doi:10.1007/s10126-006-6045-1)
114. Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, Wilkinson GS. 2015 Cancer across the tree of life: cooperation and cheating in multicellularity. *Phil. Trans. R. Soc. Lond. B* **370**, 20140219. (doi:10.1098/rstb.2014.0219)
115. Lawrence MS *et al.* 2013 Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218. (doi:10.1038/nature12213)
116. Bachère E, Gueguen Y, Gonzalez M, de Lorgeril J, Garnier J, Romestand B. 2004 Insights into the anti-microbial defense of marine invertebrates: the penaeid shrimps and the oyster *Crassostrea gigas*. *Immunol. Rev.* **198**, 149–168. (doi:10.1111/j.0105-2896.2004.00115.x)
117. Hoxhaj G, Manning BD. 2020 The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. *Nat. Rev. Cancer* **20**, 74–88. (doi:10.1038/s41568-019-0216-7)
118. Magaway C, Kim E, Jacinto E. 2019 Targeting mTOR and metabolism in cancer: lessons and innovations. *Cells* **8**, 1584. (doi:10.3390/cells8121584)
119. Zou Z, Tao T, Li H, Zhu X. 2020 mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. *Cell Biosci.* **10**, 31. (doi:10.1186/s13578-020-00396-1)
120. Hamidi H, Ivaska J. 2018 Every step of the way: integrins in cancer progression and metastasis. *Nat. Rev. Cancer* **18**, 533–548. (doi:10.1038/s41568-018-0038-z)
121. Suquet M, de Kermoysan G, Gonzalez Araya R, Queau I, Lebrun L, Le Souchu P, Mingant C. 2009 Anesthesia in Pacific oyster *Crassostrea gigas*. *Aquat. Living Resour.* **22**, 29–34. (doi:10.1051/alr/2009006)
122. Song L, Florea L. 2015 Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* **4**, 48. (doi:10.1186/s13742-015-0089-y)
123. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)
124. Fraïsse C, Belkhir K, Welch JJ, Bierne N. 2016 Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol. Ecol.* **25**, 269–286. (doi:10.1111/mec.13299)
125. Popovic I, Matias AMA, Bierne N, Riginos C. 2020 Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. *Evol. Appl.* **13**, 515–532. (doi:10.1111/eva.12857)

126. Simon A *et al.* 2020 Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evol. Appl.* **13**, 575–599. (doi:10.1111/eva.12879)
127. Grabherr MG *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/nbt.1883)
128. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. (doi:10.1093/bioinformatics/btv351)
129. Bairoch A, Apweiler R. 2000 The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48. (doi:10.1093/nar/28.1.45)
130. Nguyen VH, Lavenier D. 2009 PLAST: parallel local alignment search tool for database comparison. *BMC Bioinf.* **10**, 329. (doi:10.1186/1471-2105-10-329)
131. Blum M *et al.* 2021 The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**(D1), 344–354. (doi:10.1093/nar/gkaa977)
132. Huerta-Cepas J *et al.* 2019 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**(D1), 309–314. (doi:10.1093/nar/gky1085)
133. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017 Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419. (doi:10.1038/nmeth.4197)
134. Anders S, Huber W. 2010 Differential expression analysis for sequence count data. *Genome Biol.* **11**, NR106. (doi:10.1186/gb-2010-11-10-r106)
135. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. 2010 Enrichment map: a network-based method for gene-set enrichment visualisation and interpretation. *PLoS ONE* **5**, e13984. (doi:10.1371/journal.pone.0013984)
136. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. (doi:10.1101/gr.1239303)
137. Szklarczyk D *et al.* 2021 The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**(D1), 605–612. (doi:10.1093/nar/gkaa1074)
138. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017 Variant review with the Integrative Genomics Viewer (IGV). *Cancer Res.* **77**, 31–34. (doi:10.1158/0008-5472.CAN-17-0337)
139. Grubaugh ND *et al.* 2019 An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8. (doi:10.1186/s13059-018-1618-7)
140. Burioli EAV, Hammel M, Vignal E, Vidal-Dupiol J, Mitta G, Thomas F, Bierne N, Destoumieux-Garzón D, Charrière GM. 2023 Transcriptomics of mussel transmissible cancer MtrBTN2 suggests accumulation of multiple cancer traits and oncogenic pathways shared among bilaterians. Figshare. (doi:10.6084/m9.figshare.c.6873517)