



# Novel functional proteins coded by the human genome discovered in metastases of melanoma patients

Aniel Sanchez · Magdalena Kuras · Jimmy Rodriguez Murillo · Indira Pla · Krzysztof Pawlowski · A. Marcell Szasz · Jeovanis Gil · Fábio C. S. Nogueira · Yasset Perez-Riverol · Jonatan Eriksson · Roger Appelqvist · Tasso Miliotis · Yonghyo Kim · Bo Baldetorp · Christian Ingvar · Håkan Olsson · Lotta Lundgren · Henrik Ekedahl · Peter Horvatovich · Yutaka Sugihara · Charlotte Welinder · Elisabet Wieslander · Ho Jeong Kwon · Gilberto B. Domont · Johan Malm · Melinda Rezeli · Lazaro Hiram Betancourt · György Marko-Varga

Received: 21 May 2019 / Accepted: 2 September 2019 / Published online: 10 October 2019  
© The Author(s) 2019

**Abstract** In the advanced stages, malignant melanoma (MM) has a very poor prognosis. Due to tremendous efforts in cancer research over the last 10 years, and the introduction of novel therapies such as targeted therapies and immunomodulators, the rather dark horizon of the median survival has dramatically changed from

under 1 year to several years. With the advent of proteomics, deep-mining studies can reach low-abundant expression levels. The complexity of the proteome, however, still surpasses the dynamic range capabilities of current analytical techniques. Consequently, many predicted protein products with potential biological

---

Aniel Sanchez, Magdalena Kuras, Lazaro Hiram Betancourt, and György Marko-Varga should be considered joint first and last authors, respectively.

---

A. Sanchez (✉) · I. Pla · K. Pawlowski · J. Malm  
Section for Clinical Chemistry, Department of Translational Medicine, Skåne University Hospital Malmö, Lund University, 205 02 Malmö, Sweden  
e-mail: aniel.sanchez@med.lu.se

M. Kuras · J. R. Murillo · J. Gil · J. Eriksson · R. Appelqvist · Y. Kim · M. Rezeli · L. H. Betancourt (✉) · G. Marko-Varga  
Clinical Protein Science & Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University, BMC D13, 221 84 Lund, Sweden  
e-mail: lazaro.betancourt@med.lu.se

K. Pawlowski  
Biology, Warsaw University of Life Sciences, Warsaw, Poland

A. M. Szasz  
Cancer Center, Semmelweis University, Budapest 1083, Hungary

F. C. S. Nogueira · G. B. Domont  
Proteomics Unit, Department of Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

F. C. S. Nogueira  
Laboratory of Proteomics, LADETEC, Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Y. Perez-Riverol  
European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, CB10 1SD Hinxton, Cambridge, UK

T. Miliotis  
AstraZeneca R&D, Mölndal, Sweden

B. Baldetorp · H. Olsson · L. Lundgren · H. Ekedahl · Y. Sugihara · C. Welinder · E. Wieslander  
Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, 221 85 Lund, Sweden

C. Ingvar  
Department of Surgery, Clinical Sciences, Skåne University Hospital, Lund University, Lund, Sweden

functions have not yet been verified in experimental proteomic data. This category of ‘missing proteins’ (MP) is comprised of all proteins that have been predicted but are currently unverified. As part of the initiative launched in 2016 in the USA, the European Cancer Moonshot Center has performed numerous deep proteomics analyses on samples from MM patients. In this study, nine MPs were clearly identified by mass spectrometry in MM metastases. Some MPs significantly correlated with proteins that possess identical PFAM structural domains; and other MPs were significantly associated with cancer-related proteins. This is the first study to our knowledge, where unknown and novel proteins have been annotated in metastatic melanoma tumour tissue.

**Keywords** Melanoma · Missing proteins · Tissue · Biobank · Proteomics · Mass spectrometry

## Introduction

Metastatic melanoma is an aggressive disease; previously known to resist most types of therapies. However, the development of targeted therapies in tumours with BRAF mutations has revolutionised treatment. Nevertheless, a significant number of patients with BRAF V600 metastatic melanoma experience relapse within a few months after treatment with the combination of BRAF and MEK inhibitors (Pascale et al. 2018). With the advent of immunotherapy, a significant improvement in survival has become evident (Eroglu et al. 2018). Nonetheless, the disease often overcomes therapeutic blockage of the immune system.

New and promising classification systems and methods have emerged that have enabled stratification of patients into refined prognostic clusters. Such

approaches undoubtedly complement available therapies. As such, a more uniform prognosis is provided and, more importantly, an improved response to treatment (Pimiento et al. 2013; Tímár et al. 2016; Dimitriou et al. 2018). Based on genetic analyses, cutaneous melanomas are divided into four classes: BRAF-mutated, RAS-mutated, NF-1-mutated tumours, and triple wild-type (Cancer Genome Atlas Network et al. 2015). Independent of these sub-groups, immune therapy with check-point inhibitors across tumours has resulted in an improved outcome. Applying transcriptomic profiling and using paired-end massively parallel sequencing of cDNA together with analyses of high-resolution chromosomal copy number data, 11 novel melanoma gene fusion products and 12 novel readthrough transcripts have been identified. From this RNA-seq analysis, a surprisingly high mutational burden was described in melanoma that was crucial for tumour progression (Berger et al. 2010).

Heterogeneity, clonal expansion and evolutionary processes are further key phenomena that may be responsible for the resistance mechanism of cancer (Marcell Szasz et al. 2019; Turajlic et al. 2019; Swanton 2018). A deeper understanding of single individual tumour can reveal important pieces of the entire puzzle. For example, immunotherapies are now administered in earlier stages and it was shown that neoadjuvant ipilimumab + nivolumab expand more tumour-resident T cell clones than adjuvant application (Blank et al. 2018). The adverse effects have prompted further studies and approaches to apply immunotherapies in a safer manner (Bosman et al. 2010).

In order to address unsolved clinical drawbacks, alternative research approaches have emerged. Proteomics has been successfully applied to several biological scenarios as an integral part of multi-omics studies in system biology and medicine (Collins and Varmus 2015; Chen and Snyder 2013).

By nature, proteins are highly complex. Therefore, as a consequence of the dynamic range and sensitivity limits of current proteomic techniques, many predicted protein products have not yet been identified in proteomic experiments. These proteins could provide essential clues to aid interpretation of biological processes and potentially drive new avenues of research and therapeutic strategies to solve remaining clinical problems.

In 2016, the Chromosome-centric Human Proteome Project (C-HPP) launched an initiative to accelerate the identification and assignment of these ‘missing

---

L. Lundgren

Department of Hematology, Oncology and Radiation Physics,  
Skåne University Hospital, Lund, Sweden

P. Horvatovich

Department of Analytical Biochemistry, Faculty of Science and  
Engineering, University of Groningen, Groningen,  
The Netherlands

H. J. Kwon

Department of Biotechnology, Yonsei University, Seoul, South  
Korea

proteins' (MPs) (Omenn et al. 2017). The proteins were divided into five groups according to the level of protein existence (PE). PE1 contains proteins identified by mass spectrometry, 3D structure, immunohistochemistry, and/or amino acid sequencing. PE2 refers to transcript expression, but not protein expression. Proteins annotated in PE3 do not have any protein or transcript evidence in humans; however, there are similar sequences that have been reported in other species. PE4 proteins are hypothesised from gene models, and the PE5 group contains predicted protein sequences with uncertain evidence and is mostly associated with pseudogenes (Paik et al. 2018).

The samples in this study are a part of the BioMel biobank, governed by Lund Melanoma Study Group (LMSG). It is a collection of blood and tissue (primary and metastases) samples with detailed clinical information from patients diagnosed with malignant melanoma in Southern Sweden. Since 2013, the sample collection is prospective, including fresh frozen tissue and blood. Our biobank coupled high-end proteomic platform was used to study the melanoma tumour tissues (Welinder et al. 2015, 2017; Welinder et al., 2014a, b; Gil et al. 2019; Kuras et al. 2018; Murillo et al. 2018). We used histopathological characterisation and a genomic data-directed proteomic strategy to successfully identify a protein expression pattern that was associated with improved survival prognostics in lymph node samples from stage 3 malignant melanoma patients (Betancourt et al. 2019). A progression from locoregional to distantly spread disease was witnessed throughout the years (Fig. 1). In a more recent work, a deeper investigation has been undertaken to identify proteins in metastatic disease, namely, those that may be responsible for further progression (Gil et al. 2019).

As a consequence of the high diversity of individuals, it is crucial to perform large-scale analyses of clinical samples. This enables the identification of the highest number of proteins possible, including proteins that have never been previously reported by mass spectrometry.

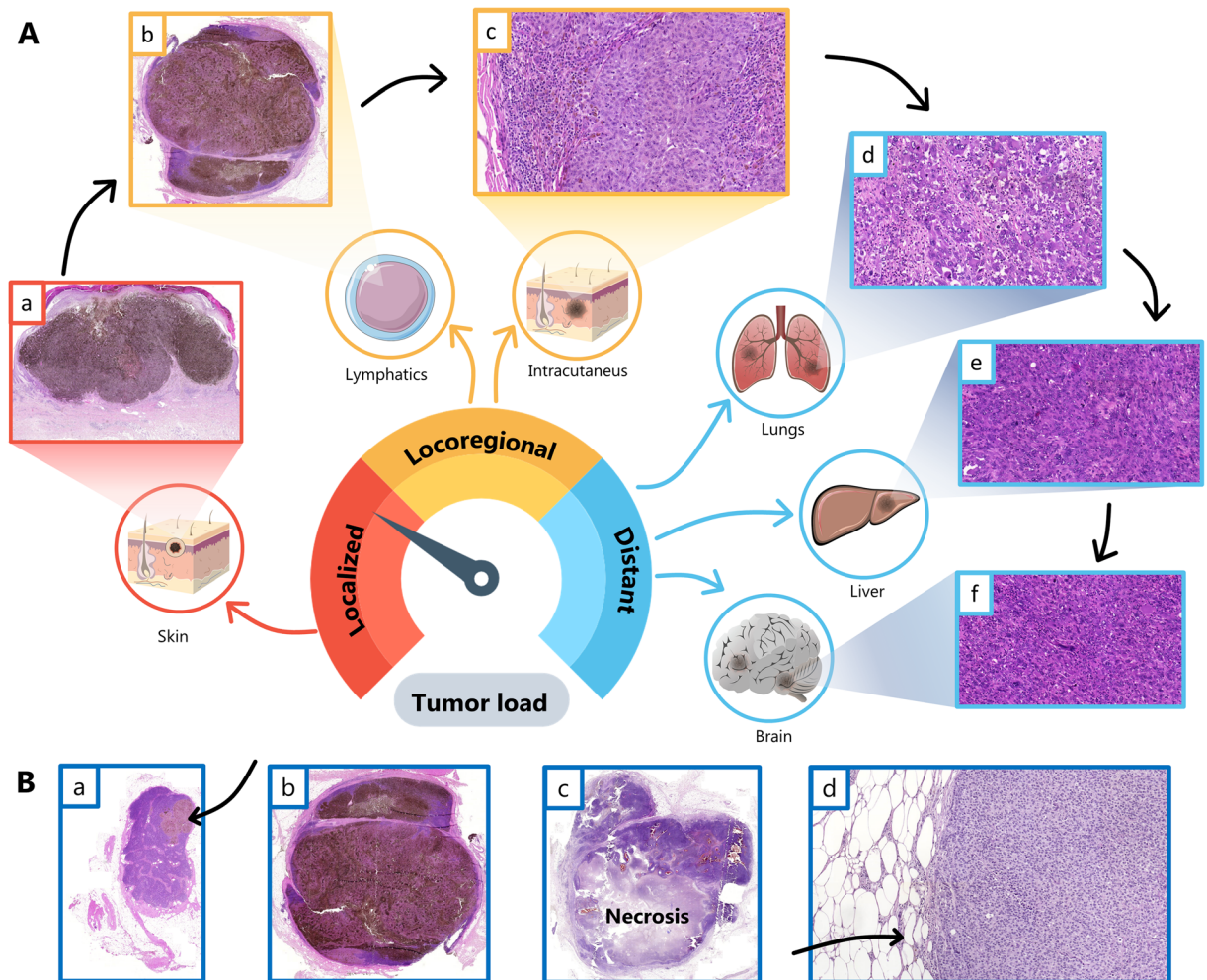
In the current study, a novel data set of 33 proteins is presented. These proteins were identified across 140 lymph node metastatic tumour samples from malignant melanoma patients. All identified proteins are currently annotated in Nextprot (Gaudet et al. 2015) as 'missing proteins'. According to the HUPO guidelines, 9 of the proteins were confidently identified by mass

spectrometry. Association clusters were constructed to pinpoint predicted functional annotations for these proteins.

## Materials and methods

This study was approved by the Regional Ethical Committee at Lund University, Southern Sweden, approval numbers: DNR 191/2007, 101/2013 and 2015/266, 2015/618. All patients involved in the study provided written informed consent. The malignant melanoma lymph node metastases were collected from patients undergoing surgical resection at Lund University Hospital, Sweden. Out of the 140 tumours included in this study, only four received any of the novel therapies. Nevertheless, the majority of the patients enrolled in the study died due to the progression of the disease. Histopathological analysis of the tissues was performed by a board-certified pathologist (Gil et al. 2019). Protein extraction and digestion were performed according to the protocol described by Kuras et al. (2018), and the resultant peptides were labelled with TMT 11-plex reagents (Thermo Fisher Scientific, San Jose, CA, USA) according to the instructions provided. Labelled peptides were separated into 24 fractions by basic reversed-phase liquid chromatography on a Phenomenex Aeris C8 column (100 mm × 2.1 mm, 3.6- $\mu$ m particles) using an Agilent 1100 HPLC system.

LC-MS/MS analysis was performed on an UltiMate 3000 RSLCnano system coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, San José, CA, USA). Data were acquired in DDA, with the ADP set to off, selecting the top 20 precursors. Full MS scans were acquired over  $m/z$  350–1400 range at a resolution of 120,000 (at  $m/z$  200), target AGC value of  $3 \times 10^6$ , maximum injection time of 50 ms, and normalised collision energy of 34%. The tandem mass spectra were acquired in the Orbitrap mass analyser with a resolution of 45,000, a target ACG value of  $1 \times 10^3$  and a maximum injection time of 86 ms. An isolation window of 0.7  $m/z$  was used and fixed first mass was set to 110  $m/z$ . Data were processed with Proteome Discoverer v2.3 (Thermo Fisher Scientific, San José, CA, USA) and searched against the *Homo sapiens* UniProt revised database (2018-10-01), including isoforms, with Sequest HT. Cysteine



**Fig. 1** (A) Life history of a melanoma. The image depicts the evolving progression of a malignant melanoma originating from the skin, spreading to the lymphatic system and giving rise to transit (intracutaneous) and distant metastases (lung, liver, and eventually brain). The histological images in chronological order: (a) primary nodular melanoma (1×, HE), (b) lymph node metastasis (1×, HE), (c) lymph node metastasis composed of epithelioid tumour cells (20×, HE), (d) lung metastasis in fibrotic background and presence of tumour infiltrating lymphocytes (20×, HE), (e)

liver metastasis—note the brisk mitotic activity and morphological change in cell shape, spindly melanoma cells (20×, HE), (f) brain metastasis of spindle and ‘monster’ melanocytes (20×, HE). (B) Metastatic melanoma in the lymphatic system in four patients. (a) Small and circumscribed melanoma in a lymph node (1×, HE). (b) Large pigmented melanoma filling the lymph node (1×, HE). (c) Large melanoma with necrotic areas (1×, HE). (d) Melanoma breaking the capsule of the lymph node and infiltrating the neighbouring tissue (20×, HE)

carbamidomethylation was set as fixed modification and methionine oxidation, protein N-terminal acetylation, TMT6plex (+ 229.163 Da) at N-terminal and lysine residues were set as dynamic modifications. Peptide mass tolerance for the precursor ions and MS/MS spectra were 10 ppm and 0.02 Da, respectively.

Protein evidence (PE) was determined using the criteria adopted from neXtProt and the Chromosome-centric Human Protein Project (C-HPP) (Omenn et al. 2018).

## Bioinformatics

### Missing protein identification

Peptide-spectrum match (PSM), peptide, and protein identifications were filtered to less than 1% FDR. Identification and sorting of unique peptides were carried using the neXtProt tool ‘Peptide uniqueness checker’ (<https://www.nextprot.org/tools/peptide-uniqueness-checker>) for all peptide sequences from proteins classified by neXtProt as P2-P5. PSMs mapping to



missing proteins were also manually inspected. All novel peptides (peptides without MS evidence) were aligned using BLASTp (version: 2.7.1) to three different databases UniProt (release date: 2018), Ensembl (release date: 2019), and RefSeq (release date: 2019) as previously suggested (Nesvizhskii 2014). All possible peptide variants were filtered using the following filters: identity score higher 70, less than 2 amino acids substitutions with respect to the original novel peptide, and theoretical mass within 10 ppm compared with the precursor mass. In addition, novel unique peptides were searched in PeptideAtlas (<http://www.peptideatlas.org>) to explore previously reported evidence in public proteomics data (Fig. 2a).

### Structural and functional identification

Structural domains in the novel proteins were identified by the conserved domain search tool (Marchler-Bauer et al. 2017). Additionally, structural domains were predicted by the FFAS and HHPred algorithms (Jaroszewski et al. 2011; Zimmermann et al. 2018).

The bioinformatics analysis of relational networks between proteins that correlated with novel PE5 proteins was performed by ingenuity pathway analysis (IPA, Qiagen, Inc., Redwood City, CA, USA). The queried data sets that were generated for the PE5 proteins were significant as assessed by adjusted  $p$  value  $< 0.01$  and included proteins with an expression correlation to a given PE5 protein across the samples in our study. Additionally, IPA provided overrepresented functional annotations and pathways within the identified subnetworks.

Protein family annotation (PFAM) of the PE2 proteins was detected using the DAVID bioinformatics database (Huang et al. 2009a, b). Spearman rank correlation test was performed to determine the correlation coefficient between PE2 proteins and protein members within the same family. The analysis was based on protein intensities that were quantitated considering unique peptides only. Correlations with  $p$  values  $< 0.05$  were considered significant.

## Results and discussions

Well-characterised samples from 140 patients with stage 3 malignant melanoma (at the time of tissue collection) were investigated. A robust workflow (Fig. 2a) was

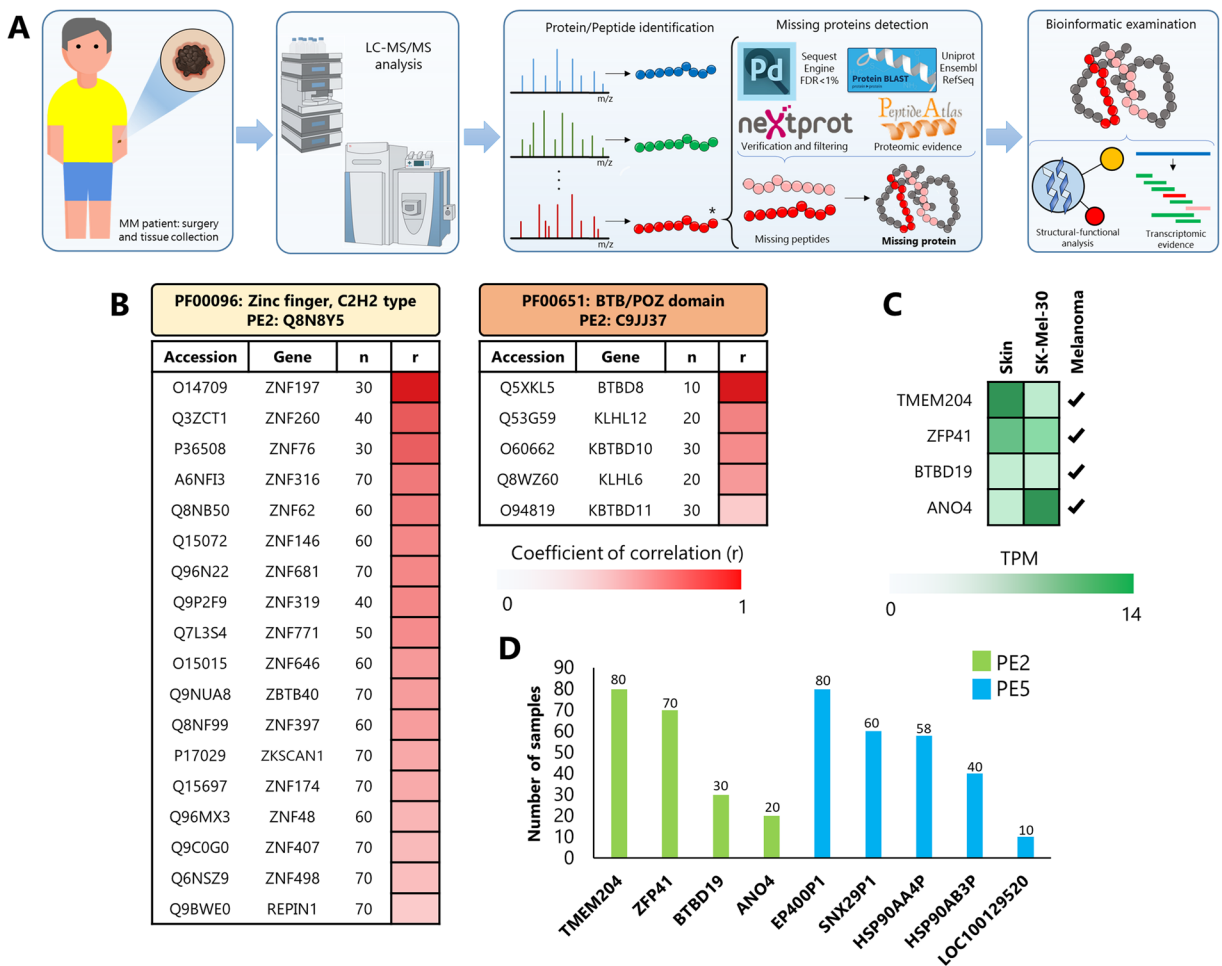
implemented that combines an automated biobank platform, advanced high-throughput proteomics, and bioinformatics. Briefly, the tissue samples were collected from patients and stored with all clinical data in a quality-controlled biobank (Welinder et al. 2013). The samples were processed with modern and reproducible proteomic techniques. To obtain all possible information related to the identified proteins, the data generated was processed with a range of bioinformatics tools.

All novel peptides were mapped to Ensembl, Refseq, and UniProt with allowance for amino acid substitutions and gaps. The aim was to determine if variants of the same peptide were apparent in other proteins and could thus explain the mass spectra. More than 5000 possible variants were returned, but none passed the criteria, i.e., a tryptic peptide with a theoretical mass  $\pm 10$  ppm of the experimental mass.

All tumours are unique in morphology and underlying biological processes; however, some drivers are shared amongst melanomas. The high number of processed heterogeneous tumour tissues enabled the identification of 33 ‘missing proteins’ across the 140 samples (Table 1). All proteins were classified according to the PE category reported by neXtProt. Annotations were applied according to the HUPO guidelines, namely, ‘two or more distinct, uniquely mapping, non-nested peptide sequences per protein of length  $\geq 9$  amino acids’ (Omenn et al. 2017). After applying these guidelines, the number of missing proteins was reduced to nine (Table 1) and they can be divided into two groups (PE2 and PE5):

1. Proteins uniquely identified in this study within the context of metastatic cancer progression: Q9BSN7 (TMEM204), Q8N8Y5 (ZFP41), C9JJ37 (BTBD19), Q32M45 (ANO4) although previously supported only by transcript presence (**PE2**)
2. Proteins where the annotation was confirmed and explicitly linked the proteins to mechanisms of melanoma metastasis: Q58FG1 (HSP90AA4P), Q6ZTU2 (EP400P1), Q8IUI4 (Putative SNX29P2), Q58FF7 (HSP90AB3P), A0A0J9YWL9 (TEX13C) while previously marked as proteins of uncertain evidence and suspected to be pseudogenes (**PE5**)

The remaining 24 proteins were identified in up to 140 melanoma metastases (Table 1). As most of the



**Fig. 2** Experimental workflow and information related to the nine ‘missing proteins’ reported. (A) A total of 140 MM tissues were analysed by LC-MS/MS. MS/MS spectra were contrasted with available databases and with annotation levels of protein identification (PE1–5). Missing proteins were evaluated in terms of peptide length, number of peptides, structural and functional analysis, and transcriptomic evidence comparison. (B) Protein spearman correlation based on the expression of two of the PE2 proteins

missing proteins are possibly low-abundance proteins (Wei et al. 2016), these results can be considered as further evidence to support the existence of the proteins in the tumours.

Expression correlation is known to be an indicator of functional association between genes or proteins (Pita-Juárez et al. 2018). Four individual Spearman correlation tests were performed to determine if there are any possible functional associations between the four PE2 proteins and well-known proteins with similarities in function, structure, or sequence. Using the protein intensities obtained from the MS data, for each novel

and proteins belong to the Zinc Finger, H2H2 type, and BTP/POZ domain. The number of samples and the  $r$  value for the spearman correlation are represented by  $n$  and  $r$  respectively. (C) Evidence-based on The Human Atlas for the four PE2 missing proteins in skin tissues, melanoma cell lines (SK-Mel-30), or melanoma tissues, TPM (Transcripts per Kilobase Million). (D) Frequency of identification across the 140 tumour samples; Y-axis represents the number of samples where the proteins were identified

protein, the correlation was assessed against proteins that have the same PFAM structural domains.

Two of the four proteins annotated as PE2 (Q8N8Y5/ZFP41 and C9JJ37/BTBD19) were significantly correlated with proteins possessing the C2H2 zinc finger domain (PF00096) and BTB/POZ domain (PF00651), respectively. Each protein was individually associated with one different protein family as shown in Fig. 2B. The Q32945/ANO4 protein was not significantly correlated with any proteins of the same family (anoctamin, calcium-activated chloride channel, PF04547); and the Q9BSN7/TMEM204 protein does not have close

**Table 1** Total list of ‘missing proteins’ identified in this study. The first nine proteins were identified with at least 2 peptides with  $\geq 9$  amino acids

No.	Protein accession <sup>a</sup>	Gene symbol	Description	Chromosomal position	Peptide sequence	No. of total PSMs <sup>b</sup>	Coverage [%] <sup>c</sup>	No. of samples detected <sup>d</sup>	neXtProt PE level
1	C9JH37	BTBD19	BTB/POZ domain-containing protein 19	1p34.1	VGAAVLERPVAEVAAPVVK QEVFAHR	1 1	15	30	PE2
2	Q32M45	ANO4	Anoctamin-4	12q23.3	LALLAPAELSALEEQNR ESSLNSDIIFVK LHAPWEVLGR ETLPDLEENDCYTAPFSQQR ISFPQWEK	1 2 1 1	6	20	PE2
3	Q8N8Y5	ZFP41	Zinc finger protein 41, homologue	8q24.3	AFNCGSNLLK EADYQK	5 3	20	70	PE2
4	Q9BSN7	TMEM204	Transmembrane protein 204	16p13.3	TEPCLSPDEDEHVDFDAFDFK GLDNDYVESPC SCWLVDR	2 1 1	23	80	PE2
5	A0A0J9YWL9	TEX13C	Putative testis-expressed protein 13C	Xq25	GGPSFGAR AGQVDAHDCEALGWGSEAGFQESR	2 6	4	10	PE5
6	Q58FF7	HSP90AAB3P	Putative heat shock protein HSP 90-beta-3	4q21-q25	EMVPLGDSHSLK SLTSDWEDHILAVK HLEINPDHPMETLR	1 3 2	36	40	PE5
7	Q58FG1	HSP90AA4P	Putative heat shock protein HSP 90-alpha A4	4q35.2	DLJMDNCEELIPEYLNFR EDLELPEDEEEK	9 2	19	58	PE5
8	Q8IU14	SNX29P2	Putative protein SNX29P2	16p11.2	ESTQNVTLTK ESTQGVSSVFR	7 3	30	60	PE5
9	Q6ZTU2-6	EP400P1	Isoform 5 of Putative EP400-like protein	12q24.33	QNLDLIEEEEEHEFVNDVK TSAAFPQQQLQVLSDSGSTVQLPR	2 7	25	80	PE5
10	Q5BKT4	ALG10	Dol-P-Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase	12p11.21	LNIPLPPTSR	15	23	110	PE2
11	Q7Z769	SLC35E3	Solute carrier family 35 member E3	12q15	AMTTPVIAIQTFYQK LSEQEGSR	6 18	10	140	PE2
12	Q8TBE1	CNIH3	Protein cornichon homologue 3	1q42.13	LDIFAPK	18	11	70	PE2
13	A1L157	TSPAN11	Tetraspanin-11	12p11.21	SPIDQCNPVHAR TLAENYQPGATQITASVDR QVPDSCCK	7 1 1	17	10	PE2
14	O14610	GNGT2	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-T2	17q21	EYVEAQAGNDPELK	8	39	70	PE2
15	O43374	RASA4	Ras GTPase-activating protein 4	7q22-q31.1	EAWMEPLQPTVTR	2	38	20	PE2
16	P18825	ADRA2C	Alpha-2C adrenergic receptor	4p16.1	AGAEGGAGDGGQAGFGAAESGALTASR	1	6	10	PE2
17	Q13304	GPR17	Uretil nucleotide/cysteiny1 leukotriene receptor	2q21	TNESSLSAK	1	2	9	PE2
18	Q5VVM6	CCDC30	Coiled-coil domain-containing protein 30	1p34.2	QHNSLLOEENIK ELELEVLK	4 1	3	40	PE2
19	Q7Z602	GPR141	Probable G protein coupled receptor 141	7p14.1	YGIHEEYNEEHCFK	1	7	10	PE2

**Table 1** (continued)

No.	Protein accession <sup>a</sup>	Gene symbol	Description	Chromosomal position	Peptide sequence	No. of total PSMs <sup>b</sup>	Coverage [%] <sup>c</sup>	No. of samples detected <sup>d</sup>	neXiProt PE level
20	Q86X67	NUDT13	Nucleoside diphosphate-linked moiety X motif 13	10q22.3	DASLSTAAQALLR HSLLELER	3 4	6	69	PE2
21	Q8IY85	EFCAB13	EF-hand calcium-binding domain-containing protein 13	17q21.32	EILEEVTK ILQSFVISEDNMVNIK	5 1	3	49	PE2
22	Q9UPC5	GPR34	Probable G protein coupled receptor 34	Xp11.4	IMYHINQNK FPNSGK YATTAR	3 2 4	9	90	PE2
23	Q9Y5J0	PCDHA13	Protocadherin alpha-13	5q31	VTVLENAFNGTLVIK	14	6	80	PE2
24	A0A0A0MT36	IGKV6D-21	Immunoglobulin kappa variable 6D-21	2p11.2	YASQISGVPSR	8	16	70	PE3
25	A0A075B6S6	IGKV2D-30	Immunoglobulin kappa variable 2D-30	2p11.2	VSNWDSGVPR	3	44	30	PE3
26	Q9BZK3	NACA4P	Putative nascent poly-peptide-associated complex subunit alpha-like protein	8q22.3	IEDLSQEAQLAAAEK	27	13	140	PE5
27	Q58FF3	HSP90B2P	Putative endoplasmic-like protein	15q26.3	EFEPFNWVK	20	24	130	PE5
28	Q96L14	CEP170P1	Cep170-like protein	4q26	EINDVAGEIDSVTSSGTAPSTTLVDR	9	58	90	PE5
29	Q9BYX7	POTEKP	Putative beta-actin-like protein 3	2q21.1	LCYVALDSEQEMAMAASSSVEK RGMMLTK	1 12	29	70	PE5
30	Q8NF67	ANKRD20A12P	Putative ankyrin repeat domain-containing protein 20A12 pseudogene	1q12	LEEIHLEQEQYK	1	11	10	PE5
31	A2A3N6	PIPSL	Putative PIPSK1A and PSMID4-like protein	10q23.33	SNPENNYGLITLDNDCEVLTLLTPDTGR	1	24	10	PE5
32	Q8IX06	REXO1L1P	Putative exonuclease GOR	8q21.2	LQFELLTQDQLK	1	4	10	PE5
33	POCG22	DHRS4L1	Putative dehydrogenase/reductase SDR family member 4-like 1	14q11.2	LGEPEDSLGINSFCLCEDASYLTGETVMVGGGTPSR	2	29	7	PE5

<sup>a</sup>Protein accession number in UniProt database<sup>b</sup>Total number of peptide-spectrum matches for the particular peptide sequence<sup>c</sup>Percentage of the protein sequence identified with unique and non-unique peptides<sup>d</sup>Number of melanoma tumour samples for which the protein was identified





2C). This evidence was provided by The Human Protein Atlas (Uhlén et al. 2016; Thul and Lindskog 2018). Taken together, the results are highly supportive of the presence of such proteins in stage 3 melanoma.

Seven of the nine proteins were quantitated in more than 30 samples and all nine in more than 10 samples. The identification frequencies of the nine PE2 and PE5 proteins during the whole analysis are shown in Fig. 2D.

Five of the nine novel proteins were annotated previously as the ‘suspect’ PE5 proteins (Table 1). Proteins annotated as PE5 typically have little to no information in the literature. Therefore, sets of proteins with expression patterns across the melanoma samples that correlated with the PE5 proteins identified in this study were queried. IPA provided functional relational subnetworks enriched in the correlated proteins. For TEX13C, IPA analysis of the correlated proteins resulted in a relational network that centred on hubs known for their involvement in cancer, such as the oestrogen receptor ESR1, SMAD3 (Tang et al. 2017), TGFB1, and ERK/MAPK kinases (Fig. 3). The proteins correlated with TEX13C are involved in cell-to-cell signalling and interaction, cellular growth and proliferation, and RNA post-transcriptional modification. For the proteins that significantly correlated with TEX13C expression, IPA generated the top three protein-protein functional relational subnetworks. TEX13C (LOC100129520) is a member of the TEX13 family that is comprised of two other members, TEX13A and TEX13B. The latter two proteins have been characterised to some extent. TEX13A is an RNA-binding protein (Nguyen et al. 2011) and the mouse homologue is a male germ cell-specific nuclear protein that may be involved in transcriptional repression (Kwon et al. 2016). This protein possesses an uncharacterised structural domain termed TEX13 and a zinc finger domain zf-RanBP (PFAM:PF00641).

Two putative HSP90 heat shock proteins, HSP90AA4P and HSP90AB3P, are close homologues of the HSP90 chaperones with well-known roles in cancer and well-established as cancer drug targets (Mbofung et al. 2017). In this study, we could establish protein-protein correlations, where 527 and 242 proteins were found to be significantly correlated with HSP90AB3P and HSP90AA4P, respectively. IPA analysis of these protein data sets yielded RNA post-transcriptional modification as the top, overrepresented functional annotation. Other overrepresented functional annotations included molecular transport and RNA

trafficking for HSP90AB3P, and protein synthesis and cell morphology for HSP90AA4P.

The EP400P1 protein is a homologue of the E1A-binding chromatin remodeller EP400, albeit containing only the EP400\_N domain with unknown function (Elsesser et al. 2019) and lacking a catalytic DEAD nuclease domain. Such an arrangement may indicate a regulatory function that is related to the longer homologue, EP400. A large number of proteins correlated with the expression of EP400P1 and the top functional annotations of the group were a cellular compromise, molecular transport, and cellular assembly and organisation.

The SNX29P2 protein is a homologue of sorting nexins involved in endosomal retromer complex function (Gallon and Cullen 2015), although the protein lacks important functional domains (the RUN domain that is probably involved in Ras-like GTPase signalling pathways and the phosphatidylinositol-3-phosphate-binding PX\_RUN domain). As such, SNX29P2 can be hypothesised as a modulator of the full-length homologue, sorting nexin-29. A very large set of proteins was observed to correlate with the expression of SNX29P2 and IPA revealed that cellular development, cellular growth and proliferation, and cell death and survival were the most common annotations amongst these proteins. Overall, the sets of proteins that had expression levels in the melanoma samples that correlated with the five novel PE5 proteins are indicative of cancer-related functions.

In conclusion, new protein evidence for nine ‘missing proteins’ is reported. These were expressed in lymph node metastases of malignant melanoma. The proteins were clearly identified across a large-scale analysis of clinical samples from melanoma patients. Furthermore, associations with cancer-related functions were obtained and discussed for all the reported proteins.

**Funding information** Open access funding provided by Lund University. This study was financially supported by the Berta Kamprad Foundation, ThermoFisher Scientific, Global, and Liconic Biobanking, and was also supported by grants from the National Research Foundation of Korea, funded by the Korean government (2015K1A1A2028365 and 2016K2A9A1A03904900) and Brain Korea 21 Plus Project, Republic of Korea, as well as the NIH/NCI International Cancer Proteogenome Consortium and the Mats and Stefan Paulsson Trust.

**Compliance with ethical standards** This study was approved by the Regional Ethical Committee at Lund University, Southern

Sweden, approval numbers: DNR 191/2007, 101/2013 and 2015/266, 2015/618. All patients involved in the study provided written informed consent.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Berger MF, Levin JZ, Vijayendran K, Sivachenko A, Adiconis X, Maguire J, et al. Integrative analysis of the melanoma transcriptome. *Genome Res.* 2010;20(4):413–27. <https://doi.org/10.1101/gr.103697.109>.
- Betancourt LH, Pawłowski K, Jonatan E, Marcell Szasz A, Mitra S, Pla I, et al. Improved survival prognostication of node-positive malignant melanoma patients utilizing shotgun proteomics guided by histopathological characterization and genomic data. *Sci Rep.* 2019;9(1):5154. <https://doi.org/10.1038/s41598-019-41625-z>.
- Blank CU, Rozeman EA, Fanchi LF, Sikorska K, van de Wiel B, Kvistborg P, et al. Neoadjuvant versus adjuvant ipilimumab plus nivolumab in macroscopic stage III melanoma. *Nat Med.* 2018;24(11):1655–61. <https://doi.org/10.1038/s41591-018-0198-0>.
- Bosman FT, F Carneiro, RH Hruban, and ND Theise. 2010. WHO classification of tumours of the digestive system. <https://www.cabdirect.org/cabdirect/abstract/20113051318>.
- Cancer Genome Atlas Network, Rehan A, Kadir C, Akdemir, B, Arman Aksoy, Monique Albert, Adrian Ally, Samirkumar B. Amin, Harindra Arachchi, et al. 2015. “Genomic classification of cutaneous melanoma.” *Cell* 161 (7): 1681–1696. <https://doi.org/10.1016/j.cell.2015.05.044>.
- Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, et al. Zinc-finger proteins in health and disease. *Cell Death Dis.* 2017;3(1):17071. <https://doi.org/10.1038/cddiscovery.2017.71>.
- Chen R, Snyder M. Promise of personalized omics to precision medicine. *Wiley Interdiscip Rev Syst Biol Med.* 2013;5(1): 73–82. <https://doi.org/10.1002/wsbm.1198>.
- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372(9):793–5. <https://doi.org/10.1056/NEJMp1500523>.
- Dimitriou F, Krattinger R, Ramelyte E, Barysch MJ, Micaletto S, Dummer R, et al. The world of melanoma: epidemiologic, genetic, and anatomic differences of melanoma across the globe. *Curr Oncol Rep.* 2018;20. <https://doi.org/10.1007/s11912-018-0732-8>.
- Elsesser O, Fröb F, Küspert M, Tamm ER, Fujii T, Fukunaga R, et al. Chromatin remodeler Ep400 ensures oligodendrocyte survival and is required for myelination in the vertebrate central nervous system. *Nucleic Acids Res.* 2019;47:6208–24. <https://doi.org/10.1093/nar/gkz376>.
- Eroglu Z, Zaretsky JM, Hu-Lieskovan S, Kim DW, Algazi A, Johnson DB, et al. High response rate to PD-1 blockade in desmoplastic melanomas. *Nature.* 2018;553(7688):347–50. <https://doi.org/10.1038/nature25187>.
- Gallon M, Cullen PJ. Retromer and sorting nexins in endosomal sorting. *Biochem Soc Trans.* 2015;43(1):33–47. <https://doi.org/10.1042/BST20140290>.
- Gaudet P, Michel P-A, Zahn-Zabal M, Cusin I, Duek PD, Evalet O, et al. The NeXtProt knowledgebase on human proteins: current status. *Nucleic Acids Res.* 2015;43(D1):D764–70. <https://doi.org/10.1093/nar/gku1178>.
- Gil J, Betancourt LH, Pla I, Sanchez A, Appelqvist R, Miliotis T, et al. Clinical protein science in translational medicine targeting malignant melanoma. *Cell Biol Toxicol.* 2019. <https://doi.org/10.1007/s10565-019-09468-6>.
- Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009a;37(1):1–13. <https://doi.org/10.1093/nar/gkn923>.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009b;4(1):44–57. <https://doi.org/10.1038/nprot.2008.211>.
- Jarozewski L, Li Z, Cai X-h, Weber C, Godzik A. FFAS server: novel features and applications. *Nucleic Acids Res.* 2011;39(suppl):W38–44. <https://doi.org/10.1093/nar/gkr441>.
- Kuras M, Betancourt LH, Rezel M, Rodriguez J, Szasz M, Zhou Q, et al. Assessing automated sample preparation technologies for high-throughput proteomics of frozen well characterized tissues from swedish biobanks. *J Proteome Res.* 2018;18(1):acs.jproteome.8b00792. <https://doi.org/10.1021/acs.jproteome.8b00792>.
- Kwon JT, Jin S, Choi H, Kim JJ, Jeong J, Kim JJ, et al. TEX13 is a novel male germ cell-specific nuclear protein potentially involved in transcriptional repression. *FEBS Lett.* 2016;590(20):3526–37. <https://doi.org/10.1002/1873-3468.12433>.
- Lee S-U, Maeda T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol Rev.* 2012;247(1):107–19. <https://doi.org/10.1111/j.1600-065X.2012.01116.x>.
- Lim J-H. Zinc finger and BTB domain-containing protein 3 is essential for the growth of cancer cells. *BMB Rep.* 2014a;47(7):405–10. <http://www.ncbi.nlm.nih.gov/pubmed/24856827>.
- Lim J-H. Zinc finger and BTB domain-containing protein 3 is essential for the growth of cancer cells. *BMB Rep.* 2014b;47(7):405–10. <https://doi.org/10.5483/BMBREP.2014.47.7.075>.
- Marcell Szasz A, Malm J, Rezel M, Sugihara Y, Betancourt LH, Rivas D, et al. Challenging the heterogeneity of disease presentation in malignant melanoma—impact on patient treatment. *Cell Biol Toxicol.* 2019;35(1):1–14. <https://doi.org/10.1007/s10565-018-9446-9>.
- Marchler-Bauer A, Yu B, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017;45(D1):D200–3. <https://doi.org/10.1093/nar/gkw1129>.
- Mbofung RM, McKenzie JA, Malu S, Zhang M, Peng W, Liu C, et al. HSP90 inhibition enhances cancer immunotherapy by

- upregulating interferon response genes. *Nat Commun.* 2017;8(1):451. <https://doi.org/10.1038/s41467-017-00449-z>.
- Murillo JR, Kuras M, Rezeli M, Miliotis T, Betancourt L, Marko-Varga G. Correction: automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. *PLoS One.* 2018;13(12):e0210234. <https://doi.org/10.1371/journal.pone.0210234>.
- Nakayama, Kentaro, Naomi Nakayama, Ben Davidson, Jim J-C Sheu, Natini Jinawath, Antonio Santillan, Ritu Salani, et al. 2006. “A BTB/POZ protein, NAC-1, is related to tumor recurrence and is essential for tumor growth and survival.” [www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene).
- Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods.* 2014;11(11):1114–25. <https://doi.org/10.1038/nmeth.3144>.
- Nguyen CD, Mansfield RE, Leung W, Vaz PM, Loughlin FE, Grant RP, et al. Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. *J Mol Biol.* 2011;407(2):273–83. <https://doi.org/10.1016/J.JMB.2010.12.041>.
- Omenn GS, Lane L, Lundberg EK, Overall CM, Deutsch EW. Progress on the HUPO draft human proteome: 2017 metrics of the human proteome project. *J Proteome Res.* 2017;16(12):4281–7. <https://doi.org/10.1021/acs.jproteome.7b00375>.
- Omenn GS, Lane L, Overall CM, Corrales FJ, Schwenk JM, Paik Y-K, et al. Progress on identifying and characterizing the human proteome: 2018 metrics from the HUPO human proteome project. *J Proteome Res.* 2018;17(12):4031–41. <https://doi.org/10.1021/acs.jproteome.8b00441>.
- Paik Y-K, Overall CM, Corrales F, Deutsch EW, Lane L, Omenn GS. Toward completion of the human proteome parts list: progress uncovering proteins that are missing or have unknown function and developing analytical methods. *J Proteome Res.* 2018;17(12):4023–30. <https://doi.org/10.1021/acs.jproteome.8b00885>.
- Pascale F, Dummer R, Paolo A Ascierio, Helen J Gogas, Ana Arance, Mario Mandala, Gabriella Liszkay, et al. Encorafenib plus Binimetinib versus Vemurafenib or Encorafenib in patients with BRAF-mutant melanoma (COLUMBUS): a multicentre, open-label, randomised phase 3 trial. *Articles Lancet Oncol.* 2018;19:603–18. [https://doi.org/10.1016/S1470-2045\(18\)30142-6](https://doi.org/10.1016/S1470-2045(18)30142-6).
- Pimiento JM, Larkin EM, Smalley KSM, Wiersma GL, Monks NR, Fedorenko IV, et al. Melanoma genotypes and phenotypes get personal. *Lab Investig.* 2013;93(8):858–67. <https://doi.org/10.1038/labinvest.2013.84>.
- Pita-Juárez Y, Altschuler G, Kariotis S, Wei W, Koler K, Green C, et al. The pathway coexpression network: revealing pathway relationships. *PLoS Comput Biol.* 2018;14(3). <https://doi.org/10.1371/journal.pcbi.1006042>.
- Swanton, Charles. 2018. “Cancer therapeutics through an evolutionary lens.” *J R Soc Med* 111 (1): 8–14. <https://doi.org/10.1177/0141076817742096>.
- Tang PM-K, Zhou S, Meng X-M, Wang Q-M, Li C-J, Lian G-Y, et al. Smad3 promotes cancer progression by inhibiting E4BP4-mediated NK cell development. *Nat Commun.* 2017;8(1):14677. <https://doi.org/10.1038/ncomms14677>.
- Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Science : A Publication of the Protein Society.* 2018;27(1):233–44. <https://doi.org/10.1002/pro.3307>.
- Timár J, Vizkeleti L, Doma V, Barbai T, Rásó E. Genetic progression of malignant melanoma. *Cancer Metastasis Rev.* 2016;35(1):93–107. <https://doi.org/10.1007/s10555-016-9613-5>.
- Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. *Nat Rev Genet.* 2019;20:404–16. <https://doi.org/10.1038/s41576-019-0114-6>.
- Uhlén M, Hallström BM, Lindskog C, Mardinoglu A, Pontén F, Nielsen J. Transcriptomics resources of human tissues and organs. *Mol Syst Biol.* 2016;12(4):862. <https://doi.org/10.15252/msb.20155865>.
- Wei W, Luo W, Wu F, Peng X, Zhang Y, Zhang M, et al. Deep coverage proteomics identifies more low-abundance missing proteins in human testis tissue with Q-exactive HF mass spectrometer. *J Proteome Res.* 2016;15(11):3988–97. <https://doi.org/10.1021/acs.jproteome.6b00390>.
- Welinder, Charlotte, Göran B. Jönsson, Christian Ingvar, Lotta Lundgren, Bo Baldetorp, Håkan Olsson, Thomas Breslin, Melinda Rezeli, Bo Jansson, Thomas E. Fehniger, et al. 2014a. “Analysis of alpha-synuclein in malignant melanoma – development of a SRM quantification assay.” Edited by Benjamin Edward Rich. *PLoS One* 9 (10): e110804. <https://doi.org/10.1371/journal.pone.0110804>.
- Welinder C, Jönsson G, Ingvar C, Lundgren L, Baldetorp B, Olsson H, et al. Feasibility study on measuring selected proteins in malignant melanoma tissue by SRM quantification. *J Proteome Res.* 2014b;13(3):1315–26. <https://doi.org/10.1021/pr400876p>.
- Welinder C, Jönsson G, Ingvar C, Lundgren L, Olsson H, Breslin T, et al. Establishing a Southern Swedish malignant melanoma OMICS and biobank clinical capability. *Clin Transl Med.* 2013;2(1):7. <https://doi.org/10.1186/2001-1326-2-7>.
- Welinder, Charlotte, Krzysztof Pawłowski, Yutaka Sugihara, Maria Yakovleva, Göran Jönsson, Christian Ingvar, Lotta Lundgren, et al. 2015. “A protein deep sequencing evaluation of metastatic melanoma tissues.” Edited by Yiqun G. Shellman. *PLoS One* 10 (4): e0123661. <https://doi.org/10.1371/journal.pone.0123661>.
- Welinder C, Pawłowski K, Szasz AM, Yakovleva M, Sugihara Y, Malm J, et al. “Correlation of histopathologic characteristics to protein expression and function in malignant melanoma.” Edited by Michal Zmijewski. *PLoS One.* 2017;12(4): e0176167. <https://doi.org/10.1371/journal.pone.0176167>.
- Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol.* 2018;430(15):2237–43. <https://doi.org/10.1016/j.jmb.2017.12.007>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.