Novel functional proteins coded by the human genome discovered in metastases of melanoma patients
Sanchez, Aniel; Kuras, Magdalena; Murillo, Jimmy Rodriguez; Pla, Indira; Pawlowski, Krzysztof; Szasz, A Marcell; Gil, Jeovanis; Nogueira, Fábio C S; Perez-Riverol, Yasset; Eriksson, Jonatan
Published in:
Cell biology and toxicology

DOI:
10.1007/s10565-019-09494-4

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
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 • Jeovani Gil • Fábio C. S. Nogueira • Yasset Perez-Riverol • Jonatan Eriksson • Roger Appelqvist • Tasso Miliotis • Yonghyo Kim • Bo Bälde torp • 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
© 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
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; Timár 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
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-μm 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 × 10⁶, 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 × 10³ and a maximum injection time of 86 ms. An isolation window of 0.7 m/z was used and fixed first mass was set to110 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
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 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)
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 than 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 ± 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 theHUPO guidelines, namely, ‘two or more distinct, uniquely mapping, non-nested peptide sequences per protein of length ≥ 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), Q8IU14 (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
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 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 C9J37/ 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 ≥ 9 amino acids

<table>
<thead>
<tr>
<th>No.</th>
<th>Protein accessiona</th>
<th>Gene symbol</th>
<th>Description</th>
<th>Chromosomal position</th>
<th>Peptide sequence</th>
<th>No. of total PSMs</th>
<th>Coverage [%]</th>
<th>No. of samples detectedb</th>
<th>neXtProt PE level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C9J357</td>
<td>BTBD19</td>
<td>BTB/POZ domain-containing protein 19</td>
<td>1p34.1</td>
<td>VVAVLFPVLEAAAPVVK 1 15 30 PE2</td>
<td>1</td>
<td>15</td>
<td>30</td>
<td>PE2</td>
</tr>
<tr>
<td>2</td>
<td>Q32M45</td>
<td>ANO4</td>
<td>Anoctamin-4</td>
<td>12q23.3</td>
<td>ESSLINSDFIFLK 1 4 6 20 PE2</td>
<td>1</td>
<td>6</td>
<td>20</td>
<td>PE2</td>
</tr>
<tr>
<td>3</td>
<td>Q8N8Y5</td>
<td>ZFP41</td>
<td>Zinc finger protein 41 homologue</td>
<td>8q24.3</td>
<td>AFNCGSLNLK EEADVQK 1 5 20 70 PE2</td>
<td>1</td>
<td>20</td>
<td>70</td>
<td>PE2</td>
</tr>
<tr>
<td>4</td>
<td>Q9BSN7</td>
<td>TMEM204</td>
<td>Transmembrane protein 204</td>
<td>1p13.3</td>
<td>GLDNVYVESC SCWLDR 1 1 23</td>
<td>1 23</td>
<td>80</td>
<td>PE2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A0A0J9YWL9</td>
<td>TEX13C</td>
<td>Putative testis-expressed protein 13C</td>
<td>Xq25</td>
<td>SRFPWEVEDR EMVPLGDSHSLK 1 4 10 PE5</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>PE5</td>
</tr>
<tr>
<td>6</td>
<td>Q58FF7</td>
<td>HSP90AB3P</td>
<td>Putative heat shock protein HSP 90-beta-3</td>
<td>4q21-q25</td>
<td>SLTSWEDHLAVK HLENPDHIPIMTLR 2 3 36</td>
<td>1 36</td>
<td>40</td>
<td>PE5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Q58FG1</td>
<td>HSP90AA4P</td>
<td>Putative heat shock protein HSP 90-alpha A4</td>
<td>4q35.2</td>
<td>DLMDNCELIPEYNFIR EDLELPDEEEEK 2 9 19 58</td>
<td>1 9</td>
<td>19</td>
<td>58</td>
<td>PE5</td>
</tr>
<tr>
<td>8</td>
<td>Q8IU4</td>
<td>SNX29P2</td>
<td>Putative protein SNX29P2</td>
<td>1p11.2</td>
<td>ESTQNVTLK ESTQGVSSVFR 1 7 30</td>
<td>1 7</td>
<td>30</td>
<td>60</td>
<td>PE5</td>
</tr>
<tr>
<td>9</td>
<td>Q6ZTU2-6</td>
<td>EP400P1</td>
<td>Isoform 5 of Putative EPI400-like protein</td>
<td>12q43.3</td>
<td>QNDLDIEEEEEEHLFEVINDEVK TSSAFAQPQPLQVLSQGTVQLPR 2 2 25</td>
<td>1 2</td>
<td>25</td>
<td>80</td>
<td>PE5</td>
</tr>
<tr>
<td>10</td>
<td>QSBK74</td>
<td>ALG10</td>
<td>Dol-1-P-Glc:Glc2Man9GlcNAc2-PP-Dol alpha-1,2-glucosyltransferase</td>
<td>12q11.21</td>
<td>LMPLPPTISR 15 12 23</td>
<td>15 12</td>
<td>23</td>
<td>110</td>
<td>PE2</td>
</tr>
<tr>
<td>11</td>
<td>QTZ769</td>
<td>SLC35E3</td>
<td>Solute carrier family 35 member E3</td>
<td>12q15</td>
<td>AMTTPVIAQTFCYQK LSEQEGSR 18 6 10</td>
<td>6 10</td>
<td>140</td>
<td>PE2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Q8TBE1</td>
<td>CNHH3</td>
<td>Protein cornichon homologue 3</td>
<td>1q22.13</td>
<td>SPIDQCPNVHAR 7 7 11</td>
<td>7 7</td>
<td>11</td>
<td>70</td>
<td>PE2</td>
</tr>
<tr>
<td>13</td>
<td>AIL157</td>
<td>TSPAN11</td>
<td>Tetraspanin-11</td>
<td>12p11.21</td>
<td>TLAYENYGQGGATQTASVR QVPDSCK 1 1 17</td>
<td>1 1</td>
<td>17</td>
<td>10</td>
<td>PE2</td>
</tr>
<tr>
<td>14</td>
<td>O14610</td>
<td>GNGT2</td>
<td>Guanine nucleotide-binding protein G1/Gy/Gso(Go) subunit gamma-T2</td>
<td>17q21</td>
<td>EYVEAQAOGNDPFLK 8 8</td>
<td>8 8</td>
<td>39</td>
<td>70</td>
<td>PE2</td>
</tr>
<tr>
<td>15</td>
<td>O43394</td>
<td>RASA4</td>
<td>Ras GTPase-activating protein 4</td>
<td>7q22-q31.1</td>
<td>EAWMEPLQPTVR 2 2 38</td>
<td>2 2</td>
<td>38</td>
<td>20</td>
<td>PE2</td>
</tr>
<tr>
<td>16</td>
<td>P18825</td>
<td>ADRA2C</td>
<td>Alpha-2C adrenergic receptor</td>
<td>4p16.1</td>
<td>AGAEQGAGGAGDQGQGAEGPAAESGALTASR 1 1 6</td>
<td>1 1</td>
<td>6</td>
<td>10</td>
<td>PE2</td>
</tr>
<tr>
<td>17</td>
<td>Q13304</td>
<td>GPR17</td>
<td>Uracil nucleotide/cysteinyl leukotriene receptor</td>
<td>2q11</td>
<td>TNESSLSAK 1 1 2</td>
<td>1 1</td>
<td>2</td>
<td>9</td>
<td>PE2</td>
</tr>
<tr>
<td>18</td>
<td>Q5VVM6</td>
<td>CCDC30</td>
<td>Coiled-coil domain-containing protein 30</td>
<td>1p34.2</td>
<td>QHSLLQEMNIK ELEELVRL 1 4 3</td>
<td>4 3</td>
<td>40</td>
<td>PE2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>QTZ602</td>
<td>GPR141</td>
<td>Probable G protein coupled receptor 141</td>
<td>7p14.1</td>
<td>YGIHEEYNEHCFK 1 1 7</td>
<td>1 1</td>
<td>7</td>
<td>10</td>
<td>PE2</td>
</tr>
<tr>
<td>No.</td>
<td>Protein accession</td>
<td>Gene symbol</td>
<td>Description</td>
<td>Chromosomal position</td>
<td>Peptide sequence</td>
<td>No. of total PSMs</td>
<td>Coverage [%]</td>
<td>No. of samples detected</td>
<td>neXtProt PE level</td>
</tr>
<tr>
<td>-----</td>
<td>------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>20</td>
<td>Q86X67</td>
<td>NUDT13</td>
<td>Nucleoside diphosphate-linked moiety X motif 13</td>
<td>10q22.3</td>
<td>DACLLSTAQLALLR</td>
<td>3</td>
<td>6</td>
<td>69</td>
<td>PE2</td>
</tr>
<tr>
<td>21</td>
<td>Q8IY85</td>
<td>EFCAB13</td>
<td>EF-hand calcium-binding domain-containing protein 13</td>
<td>17q21.32</td>
<td>EILIEVTIKNV</td>
<td>5</td>
<td>3</td>
<td>49</td>
<td>PE2</td>
</tr>
<tr>
<td>22</td>
<td>Q9UC5</td>
<td>GPR34</td>
<td>Probable G protein coupled receptor 34</td>
<td>Xp11.4</td>
<td>IMYHINQNK</td>
<td>3</td>
<td>9</td>
<td>90</td>
<td>PE2</td>
</tr>
<tr>
<td>23</td>
<td>Q9Y5I0</td>
<td>PCDHA13</td>
<td>Protocadherin alpha-13</td>
<td>5q31</td>
<td>VTVELNAPNGTLVIK</td>
<td>14</td>
<td>6</td>
<td>80</td>
<td>PE2</td>
</tr>
<tr>
<td>24</td>
<td>A0A040MT36</td>
<td>IGKV6D-21</td>
<td>Immunoglobulin kappa variable 6D-21</td>
<td>2p11.2</td>
<td>YASQISGVPDR</td>
<td>8</td>
<td>16</td>
<td>70</td>
<td>PE3</td>
</tr>
<tr>
<td>25</td>
<td>A0A05B6S66</td>
<td>IGKV2D-30</td>
<td>Immunoglobulin kappa variable 2D-30</td>
<td>2p11.2</td>
<td>VSNWDSGVDR</td>
<td>3</td>
<td>44</td>
<td>30</td>
<td>PE3</td>
</tr>
<tr>
<td>26</td>
<td>Q9BZK3</td>
<td>NACA4P</td>
<td>Putative nascent polypeptide-associated complex subunit alpha-like protein</td>
<td>8q22.3</td>
<td>IEDLSQEALAAAEEK</td>
<td>27</td>
<td>13</td>
<td>140</td>
<td>PE5</td>
</tr>
<tr>
<td>27</td>
<td>Q58FF3</td>
<td>HSP90B2P</td>
<td>Putative endoplasmic reticulum protein</td>
<td>15q26.3</td>
<td>EFEPNPWVK</td>
<td>20</td>
<td>24</td>
<td>130</td>
<td>PE5</td>
</tr>
<tr>
<td>28</td>
<td>Q96L14</td>
<td>CEP170P1</td>
<td>Cep170-like protein</td>
<td>4q26</td>
<td>EINDVAEDETVSSTGAPSTTLVDR</td>
<td>9</td>
<td>58</td>
<td>90</td>
<td>PE5</td>
</tr>
<tr>
<td>29</td>
<td>Q9BYX7</td>
<td>POTEKP</td>
<td>Putative beta-actin-like protein 3</td>
<td>2q21.1</td>
<td>LCYVALDSFQUMAMAASSSSVEK</td>
<td>1</td>
<td>29</td>
<td>70</td>
<td>PE5</td>
</tr>
<tr>
<td>30</td>
<td>Q8NI67</td>
<td>ANKR2D20A12P</td>
<td>Putative ankyrin repeat domain-containing protein 20A12 pseudogene</td>
<td>1q12</td>
<td>LEEHILQESAOYYK</td>
<td>1</td>
<td>11</td>
<td>10</td>
<td>PE5</td>
</tr>
<tr>
<td>31</td>
<td>A2A3N6</td>
<td>PIPS4</td>
<td>Putative PIP5K1A and PSMD4-like protein</td>
<td>10q23.33</td>
<td>SNPENNVLITLDNDCEVLTTTLTDPDTGR</td>
<td>1</td>
<td>24</td>
<td>10</td>
<td>PE5</td>
</tr>
<tr>
<td>32</td>
<td>Q8IY84</td>
<td>REXO1L1P</td>
<td>Putative exonuclease GOR</td>
<td>8q21.2</td>
<td>LQFLEFLQQLQK</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>PE5</td>
</tr>
<tr>
<td>33</td>
<td>P0CG22</td>
<td>DHRS4L1</td>
<td>Putative dehydrogenase/reductase SDR family member 4-like 1</td>
<td>1q11.2</td>
<td>LGEPEDLSGIVSFCLSSEDASLYTGETVMVGGTQPSR</td>
<td>2</td>
<td>29</td>
<td>7</td>
<td>PE5</td>
</tr>
</tbody>
</table>

*a* Protein accession number in UniProt database  
*b* Total number of peptide-spectrum matches for the particular peptide sequence  
*c* Percentage of the protein sequence identified with unique and non-unique peptides  
*d* Number of melanoma tumour samples for which the protein was identified
human homologues in existing protein databases. FFAS analysis of remote sequence similarities, however, showed that TMEM204 is significantly similar to claudin-like transporters that have known roles in tight junction and in cancer.

The zinc finger domain proteins have often been related to cancer progression, including several cancer forms, such as breast cancer, gastric cancer, and melanoma (Cassandri et al. 2017; Lim 2014a). To date, however, specifically, the ZFP41 gene has never been differentially detected in any melanoma study. As this constitutes the first evidence at the protein level, future studies are necessary to relate the expression of the protein with the progression of melanoma.

Conversely, the BTB/POZ domain-containing proteins are known to be involved in several types of human cancer (Nakayama et al. 2006). The BTBD19 gene has been differentially expressed in melanoma studies (Expression Atlas codes (https://www.ebi.ac.uk/arrayexpress/experiment): E-MTAB-6214, E-MTAB-7143). As this is the first time that C9JJ37/BTBD19 has been observed on protein level, future studies should be performed to confirm the specific role of this protein in melanoma. Of note, both domains (zinc finger and BTB/POZ) are structural sections of the proteins termed ‘ZBTB’, an emerging family of transcription factors with active roles in oncogenesis (Lee and Maeda 2012; Lim 2014b).

In addition, expression data analysis of the PE2 proteins (genes: BTBD19, ANO4, ZFP41, and TMEM204) revealed that all have been previously observed in skin tissues, melanoma cell lines, or melanoma tissues (Fig. 3).

**Fig. 3** Functional relationship network for proteins correlated to TEX13C. Ingenuity pathway analysis (IPA) for the proteins significantly correlated to TEX13C expression in the melanoma samples. Three top protein-protein functional relationship subnetworks merged. Red, proteins with expression positively correlated to TEX13C. Blue, proteins negatively correlated to TEX13C. Solid lines, direct functional relationships. Dashed lines, indirect relationships.
This evidence was provided by The Human Protein Atlas (Uhlin 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 2016K2A9A1A0390490) 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


Paik Y-K, Overall CM, Corrales F, Deutsch EW, Lane L, Ommen 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.


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