



## In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer

Soulaimane Aboulouard, Maxence Wisztorski, Marie Duhamel, Philippe Saudemont, Tristan Cardon, Fabrice Narducci, Anne-Sophie Lemaire, Firas Kobeissy, Eric Leblanc, Isabelle Fournier, et al.

### ► To cite this version:

Soulaimane Aboulouard, Maxence Wisztorski, Marie Duhamel, Philippe Saudemont, Tristan Cardon, et al.. In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer. Cell Reports Medicine, 2021, Cell Reports Medicine, 2 (6), pp.100318. 10.1016/j.xcrm.2021.100318 . hal-04009565

**HAL Id: hal-04009565**

**<https://hal.univ-lille.fr/hal-04009565>**

Submitted on 1 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

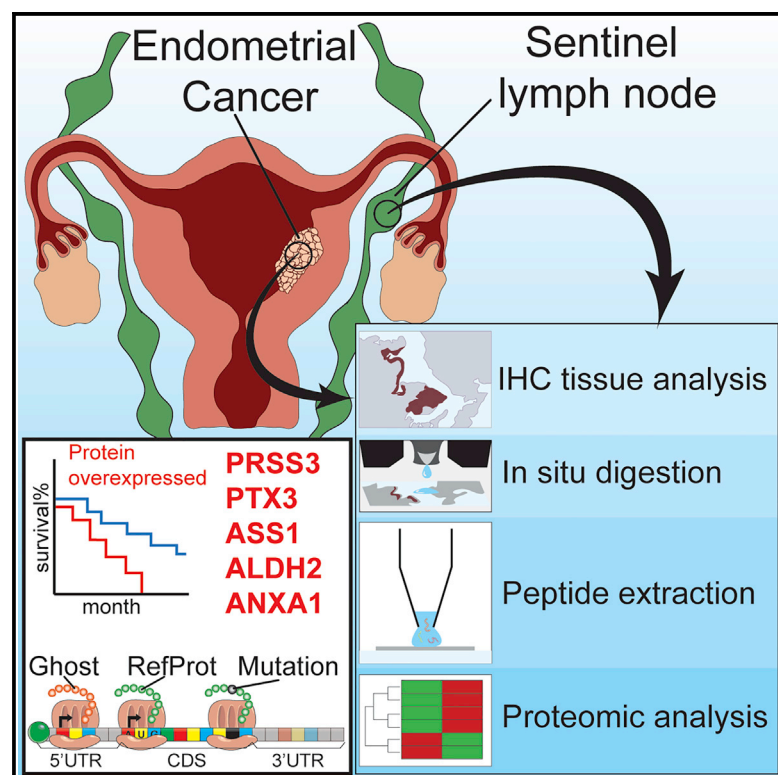
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer

## Graphical abstract



## Authors

Soulaimane Aboulouard,  
Maxence Wisztorski, Marie Duhamel, ...,  
Eric Leblanc, Isabelle Fournier,  
Michel Salzet

## Correspondence

e-leblanc@o-lambret.fr (E.L.),  
isabelle.fournier@univ-lille.fr (I.F.),  
michel.salzet@univ-lille.fr (M.S.)

## In brief

The standard of care for endometrial cancer (EC) does not currently include sentinel lymph node (SLN) mapping. Aboulouard et al. report an in-depth proteomic analysis of SLNs from individuals with grade I, II, or III EC and identify potential biomarkers for tumor grade and overall survival.

## Highlights

- Diagnosis of endometrial cancer does not include sentinel lymph node mapping
- In depth-proteomics analysis of sentinel lymph nodes and individuals with endometrial cancer
- We identify potential biomarkers for tumor grade and overall survival



## Article

# In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer

Soulaimane Aboulouard,<sup>1,4</sup> Maxence Wisztorski,<sup>1,4</sup> Marie Duhamel,<sup>1</sup> Philippe Saudemont,<sup>1</sup> Tristan Cardon,<sup>1</sup> Fabrice Narducci,<sup>1,2</sup> Anne-Sophie Lemaire,<sup>1,2</sup> Firas Kobeissy,<sup>3</sup> Eric Leblanc,<sup>1,2,\*</sup> Isabelle Fournier,<sup>1,\*</sup> and Michel Salzet<sup>1,5,\*</sup>

<sup>1</sup>Université Lille, INSERM, CHU Lille, U1192, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), 59000 Lille, France

<sup>2</sup>Department of Gynecology Oncology, Oscar Lambret Center, 59000 Lille, France

<sup>3</sup>Department of Biochemistry and Molecular Genetics, Faculty of Medicine, American University of Beirut, Beirut, Lebanon

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: [e-leblanc@o-lambret.fr](mailto:e-leblanc@o-lambret.fr) (E.L.), [isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr) (I.F.), [michel.salzet@univ-lille.fr](mailto:michel.salzet@univ-lille.fr) (M.S.)

<https://doi.org/10.1016/j.xcrm.2021.100318>

## SUMMARY

Endometrial cancer (EC) is one of the most common gynecological cancers worldwide. Sentinel lymph node (SLN) status could be a major prognostic factor in evaluation of EC, but several prospective studies need to be performed. Here we report an in-depth proteomics analysis showing significant variations in the SLN protein landscape in EC. We show that SLNs are correlated to each tumor grade, which strengthens evidence of SLN involvement in EC. A few proteins are overexpressed specifically at each EC tumor grade and in the corresponding SLN. These proteins, which are significantly variable in both locations, should be considered potential markers of overall survival. Five major proteins for EC and SLN (PRSS3, PTX3, ASS1, ALDH2, and ANXA1) were identified in large-scale proteomics and validated by immunohistochemistry. This study improves stratification and diagnosis of individuals with EC as a result of proteomics profiling of SLNs.

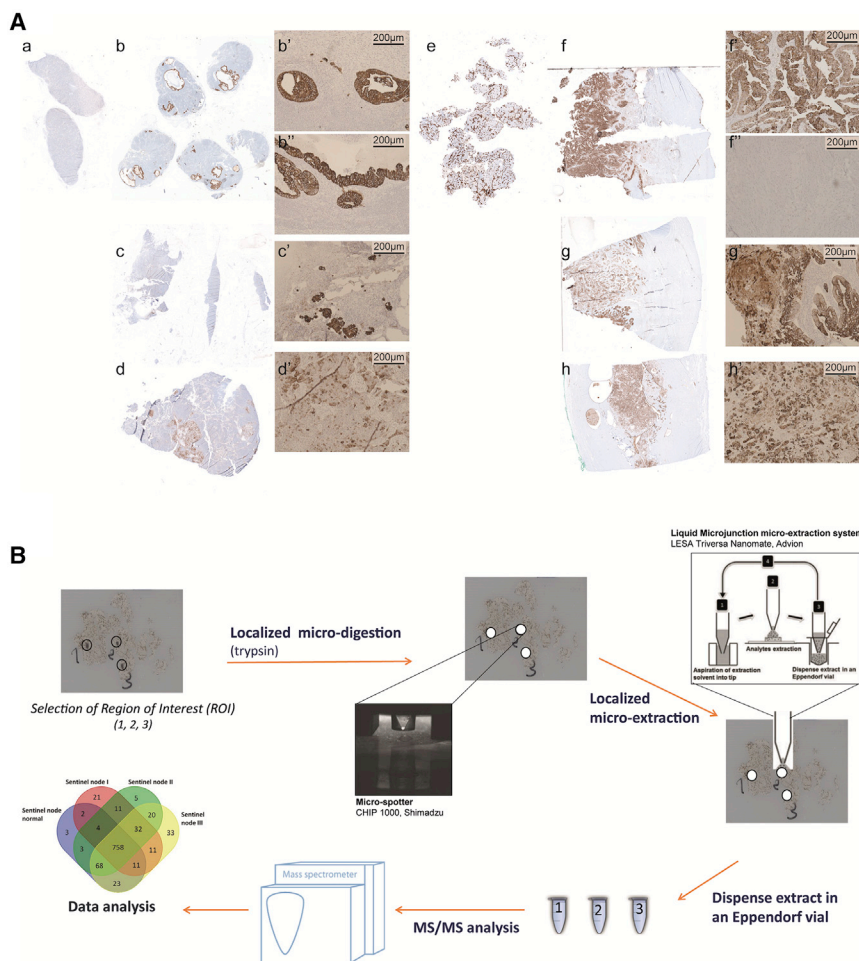
## INTRODUCTION

Sentinel lymph node (SLN) mapping is used as a surgical strategy to perform a complete lymphadenectomy in individuals with endometrial cancer.<sup>1</sup> The concept of “sentinel nodes” appeared in 1960<sup>2</sup> and is linked to the fact that if the SLNs are negative for metastasis, then nodes distal from the SLNs should also be negative.<sup>3</sup> In 1977, Cabanas<sup>4</sup> used lymphography to describe SLNs in individuals with penile carcinoma. SLN mapping enables affected individuals to avoid the side effects associated with complete lymphadenectomy and guides surgeons in decision-making. The use of pathologic “ultrastaging” and surgeon experience are key factors for successful SLN mapping, especially with breast cancer and melanoma.<sup>1,5</sup> The approach is based on simply identifying the anatomical location of the SLNs.<sup>6</sup> In the case of gynecological malignancies, the reliability of the SLN detection procedure has been investigated extensively in vulvar and cervical cancer.<sup>7</sup> SLN mapping in endometrial cancer (EC) was introduced by Burke et al.<sup>8</sup> and gained credibility in recent years<sup>9</sup> but has not yet been incorporated as a standard-of-care procedure in EC.<sup>10</sup> There are several factors that can explain the low use of SLN mapping, including complex uterine drainage, the various modalities of tracer injection, and lack of large prospective series. It has been reported that SLN mapping achieved a detection rate of 81.7%, a metastatic SLN involvement rate of

10.9%, and a false negative rate of 12.3% in main clinical trials.<sup>10</sup> In 2017, 55 eligible studies that included 4,915 women, were published.<sup>11,12</sup> The overall detection rate of SLN mapping was 81%, with a bilateral pelvic node detection rate of 50% and para-aortic detection rate of 17% with a metastasis detection sensitivity of 96%. Thus, SLN mapping accurately predicts nodal status in women with EC.<sup>11</sup> Despite these promising clinical data and a positive view of the role of SLN mapping in detecting EC, the exact underlying molecular mechanisms relating SLNs and EC grades have not been fully identified. Obtaining molecular information can highlight EC pathological mechanisms and serve to identify potential prognostic and therapeutic targets.

Here we present a state-of-the-art proteomics study of SLNs from individuals with EC to look at differences in protein expression and mutations that are associated with tumor grade and identify affected pathways possibly involved in EC. In addition to the mutations, the unreference proteins translated from the regions described as non-coding of mRNA such as 5' and 3' UTRs and frameshift or from non-coding RNA (ncRNA), forming the “ghost proteome,” were also investigated. Our work identifies a correlation between SLNs and EC grades based on significant protein abundance variation. Furthermore, we identify and validate five key protein biomarkers that link EC and SLN cancer grading that could be used later as diagnostic tools, pending validation in larger and independent cohorts from multicenters.





**Figure 1. Global workflow of spatially resolved analysis**

(A) Histopathological data obtained from immunocytochemistry performed with anti-P53 on tissue sections of sentinel nodes (SNs), either healthy (Aa) or cancerous (Ab, Ac, and Ad), from grade I (Ab) with magnification (Ab' and Ab''), grade II (Ac) with magnification (Ac'), and grade III (Ad) with magnification (Ad'). Similarly, results were obtained with their corresponding endometrial tissue, either healthy (Ae) or cancerous (Af, Ag, and Ah), at grade I (Af) with magnification (Af' and Af''), grade II (Ag) with magnification (Ag'), and grade III (Ah) with magnification (Ah'). See also IHC images in the [Supplemental information](#).

(B) Workflow for spatially resolved proteomics using IHC tissue sections. ROIs were subjected to enzymatic microdigestion using trypsin followed by liquid junction microextraction, and then subjected to shotgun proteomics analyses.

extraction before separation with a nano-liquid chromatography (LC) coupled with high-resolution mass spectrometry for tandem mass spectrometry (MS/MS) analysis. We first analyzed proteins from sentinel nodes samples and endometrioid tissue samples separately, and then we compared their protein profiles together to assess common signatures that might exist between them. In this way, we expected to demonstrate whether SLNs could provide a more sensitive method of assessing the spread of apparent early-stage EC than a lymph node dissection, which would enable a focus-targeted adjuvant therapy decision, such as performing radiotherapy or chemotherapy.

### Shotgun proteomics of sentinel nodes

From the 24 samples (normal and grades I–III), 1,291 proteins showed a significant difference in expression based on MS-based relative quantification. After filtering proteins based on a minimum number of values in at least one of the four defined groups (2 of 3 valid values), 1,005 proteins were obtained (Figure 1B; Data S1). Sixty-two specific proteins were identified with the following repartition: 3 in the normal sentinel node, 21 in grade I, 5 in grade II, and 33 in grade III (Figure 1B; Table S2). Comparison of the different sentinel node cancerous states showed differences in terms of cellular components and molecular functions. Grade I tumors contained the highest level of proteins related to the nucleus and cytoplasmic constituents compared with grade II and grade III tumors (Figure S1A), whereas the three grades presented the same level of proteins as detected in exosomes (Figure S1A). For the molecular functions, proteins identified in grade I tumors were more related to an immune response, which was confirmed by STRING analysis

## RESULTS

The aim of this study was to identify comprehensive molecular proteomics signatures from SLNs and compare them with early-stage endometrial carcinoma proteomes at intermediate and high risk of recurrence, which was performed as a randomized study (Table S1). All sentinel nodes were examined by standard staining. When negative, serial sections were performed for standard staining and pancytokeratin immunohistochemistry (IHC). For the endometrial tissues, P53 immunostaining was performed. Figure 1A presents results of IHC from a healthy sentinel node (Figure 1Aa) or normal endometrial tissue (Figure 1Ae), including from sentinel nodes from grade I (Figures 1Ab–1Ab''), grade II (Figures 1Ac and 1Ac'), and grade III (Figures 1Ad and 1Ad') or endometrioid grade I (Figures 1Af and 1Af''), grade II (Figures 1Ag and 1Ag'), and grade III (Figures 1Ah and 1Ah'). Based on these IHC results, we performed spatially resolved shotgun proteomics on regions of interest (ROIs) selected by a pathologist (Figure 1B). The stained slides were then unmounted and resin was removed, and then trypsin was deposited on the ROI using a piezo chemical inkjet printer. The digested ROI was then subjected to liquid junction peptide

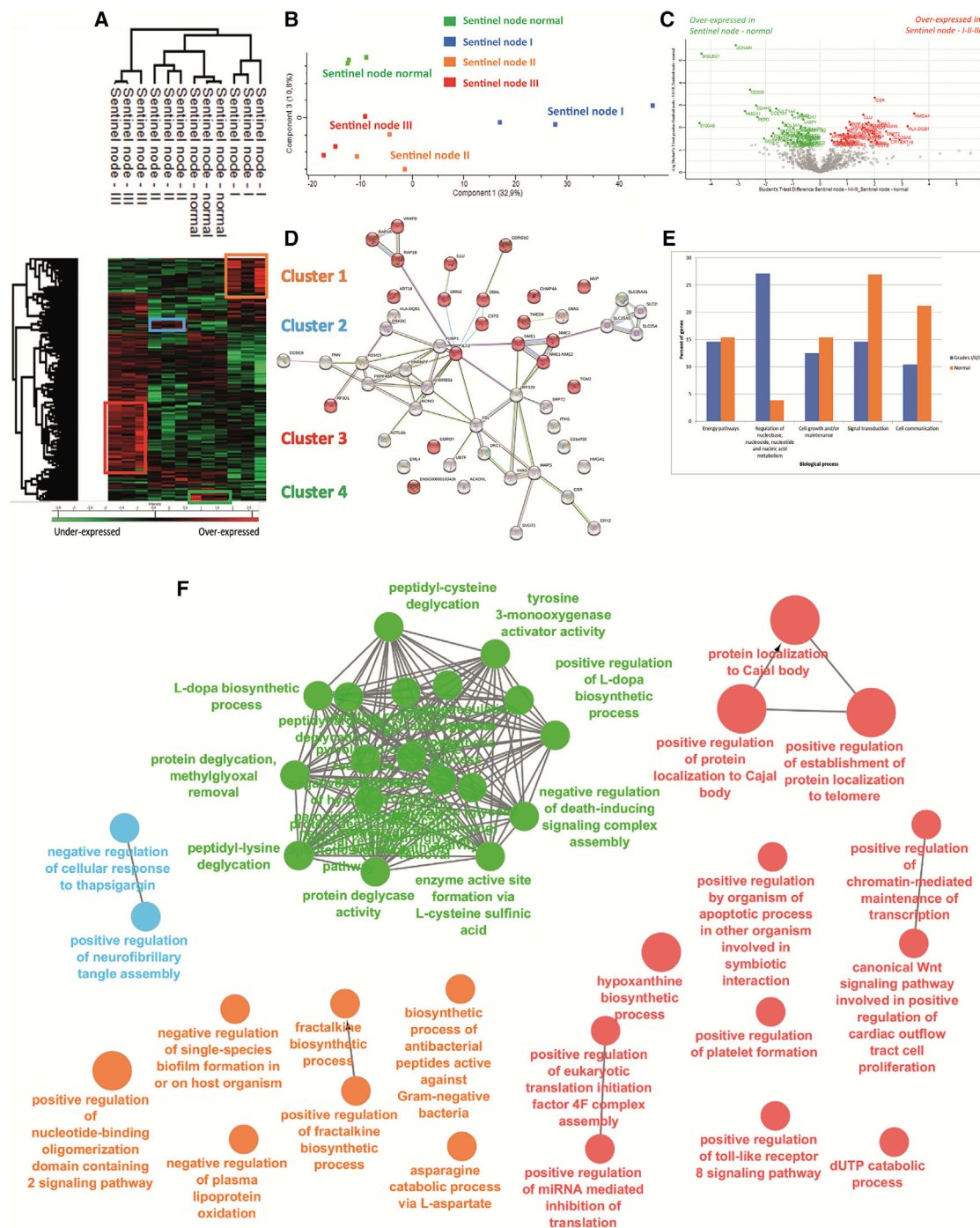


(Figure S1B). Serpins, PTX3 (pentraxin-related protein 3), CHI3L1 (chitinase-3-like protein 1), PROM1 (prominin-1), ORM2 (alpha-1-acid glycoprotein 2), MNDA (myeloid cell nuclear differentiation antigen), AZU1 (azurocidin), RNASE3 (eosinophil cationic protein), ASRGL1 (isoaspartyl peptidase/L-asparaginase), and MMP8 (neutrophil collagenase) are proteins involved in the immune response, especially the innate immune response (Figure S1B). Grade II tumors contained only 5 specific proteins: PLEK (Pleckstrin), PLIN1 (Perilipin-1), PLIN4 (Perilipin-4), ECM1 (extracellular matrix protein 1), and GIMAP1 (GTPase, IMAP family member 1). PLIN1 and PLIN4 are involved in the PPAR (peroxisome proliferator-activated receptors) signaling pathway. ECM1 and PLEK are implicated in platelet degranulation and immunity. ECM1 is also involved in angiogenesis (Figure S1C). Grade III tumors contained proteins involved in DNA and RNA binding, nucleic acid metabolism, transcription regulator activity, and metabolism (Figure S1D). Several interesting proteins can be pointed out, in particular CDC42 (cell division control protein 42 homolog), which plays a role in extension and maintenance of formation of thin, actin-rich surface projections called filopodia. SNW1 (NW domain-containing protein 1) is known to be implicated in epigenetics and is involved in NOTCH1-mediated transcriptional activation. Metastasis-associated protein (MTA2) is associated with the estrogen receptor in breast cancer and predicts proliferation in non-small cell lung cancer.<sup>13</sup> MTA2 also targets P53. MTA1, but not MTA2, has already been identified in sentinel nodes of head and neck cancer<sup>14</sup> and breast cancer.<sup>15</sup> MTA1 expression was correlated positively with lymph node metastasis and poor survival rate in EC.<sup>16</sup> According to the TCGA (The Cancer Genome Atlas)<sup>17</sup>, a list of 786 unfavorable genes has been characterized, and among them, 20 are considered to be associated with lower overall survival when overexpressed.<sup>18</sup> From this list, PTX3 protein was identified as the only unfavorable factor in grade I.

To better understand the modulation registered across the different lesions, a multiple-sample test ANOVA with  $p < 0.05$  was performed. A total of 336 proteins showed a significant difference in expression among the 4 groups, as shown in a heatmap (Figure 2A). Proteomes in samples from the same group were similar (mean Pearson correlation, 0.92) compared with inter-group variation. The main differences were observed between normal sentinel nodes and grade I SLNs (Figure 2B). A volcano plot was made, based on combination of grades I, II, and III together, which were compared with normal sentinel nodes; the data revealed 91 differentially regulated proteins, with 44 proteins more represented in tumor tissue and 47 in normal tissue (Figure 2C; Table S3). Among the 44 identified proteins more expressed in cancerous sentinel nodes, 12 are involved in the innate immune response (protein SGT1 homolog, nucleoside diphosphate kinase B, Drebrin-like protein, alpha-1-acid glycoprotein 2, Clusterin, DNA-dependent protein kinase catalytic subunit, Ras-related protein Rap-1b, vesicle-associated membrane protein 8, Ras-related protein Rap-1A, major vault protein, interleukin enhancer-binding factor 2, and Cystatin-B), 4 are involved in necroptosis and cellular senescence (ADP/ATP translocase 1, ADP/ATP translocase 2, ADP/ATP translocase 3, ADP/ATP translocase 4, and charged multivesicular body protein 4a), and 20 are involved in mediated transport (Figure 2D).

Functional enrichment analysis established that tumors are connected to nucleotide metabolism, whereas proteins involved in normal tissue are highly related to signal transduction and cell communication (Figure 2E). The hierarchical clustering between normal grades and the different grades showed a separation into 2 branches, i.e., one separating grade III and another separating the other grades and normal grade. The second branch separated grade I SLNs to normal and grade II SLNs, and then the last subbranch separated normal from grade II SLNs (Figure 2A). Cluster 1, representing overexpressed proteins in grade I SLNs, contains 47 proteins (Data S2). Nine are antimicrobial peptides, i.e., dermicidin, S100-A8, S100-A9, eosinophil cationic protein, lysozyme C, lactotransferrin, neutrophil elastase, cathepsin G, and neutrophil gelatinase-associated lipocalin. Twenty-four proteins are involved in the innate immune response or neutrophil degranulation (Figure S2A). The other proteins are involved in energy pathways and metabolism (Figure S2A). Cluster 2, corresponding to proteins overexpressed in grade II, contains 37 proteins that are involved in metabolism involving mitochondrial enzymes, such as isocitrate dehydrogenase 1 (IDH1), serine hydroxymethyltransferase (SHMT2), 2-oxoglutarate dehydrogenase (OGDH), ADP/ATP (translocase 1–4), glycanic enzymes such as UDP-glucose 6-dehydrogenase (UGDH), sialic acid synthase (NANS, N-Acetylneuraminase Synthase), 78-kDa glucose-regulated protein (HSPA5), hypoxia upregulated protein 1 (HYOU1), endoplasmic reticulum protein (HSP90A1), and protein disulfide-isomerase A3 (Figure S2B). Other proteins are also involved in cell growth (Figure S2D). Cluster 3 contains 117 proteins more expressed in grade III SLNs (Data S2). ClueGo analyses established that 32 are involved in the ribonucleoprotein complex (Figure S2C, red balls), ribosomes (Figure S2C, blue balls), and translation (Figure S2C, green balls). Among the identified proteins, ICAM-3 is known to mediate inflammatory signaling to promote cancer cell stemness.<sup>19,20</sup> Catenin beta-1 and Stathmin are poor prognosis markers in EC.<sup>21,22</sup> TP53BP1 is also detected and is known to interact with p53 and MFN1; these two genes encode a mitochondrial membrane protein and are considered to have tumor suppressor gene functions. However, its mutation is also considered a poor prognosis marker in cancer.<sup>23</sup> Functional enrichment analyses demonstrated that proteins more represented in grade III SLNs are involved in metabolism of nucleotides, protein metabolism, and energy pathways (Figure S2D). Cluster 4 corresponds to proteins that are more represented in normal sentinel nodes; 26 cytosolic proteins have been identified (Figure S2E) and are involved in oxygen carrier activity, oxygen binding, and myosin binding. Gene set enrichment analysis (GSEA)<sup>24</sup> associated with Cytoscape<sup>25</sup> analyses of the 4 clusters is in line with the precedent analyses (Figure 2F).

Comparison of the volcano plot of the grade I, grade II, and grade III sentinel nodes (Figures 3A–3C) established a molecular transition between these grades. The volcano plot confirmed the presence of a high number of differentially regulated proteins between grade I and normal SLNs (Figure 3A). No differentially regulated proteins were detected between grade II and normal SLNs (Figure 3B), whereas few proteins were differentially regulated between grade III and normal SLNs (Figure 3C). Thus, we observed a clear shift from normal to grade I sentinel nodes that keeps immune activity through antimicrobial peptides and



**Figure 2. Proteomics analysis of SLNs**

(A) Hierarchical clustering of the most variable proteins between normal tissue and grades I–III (n = 3 for each category, ANOVA with permutation-based FDR < 0.05).

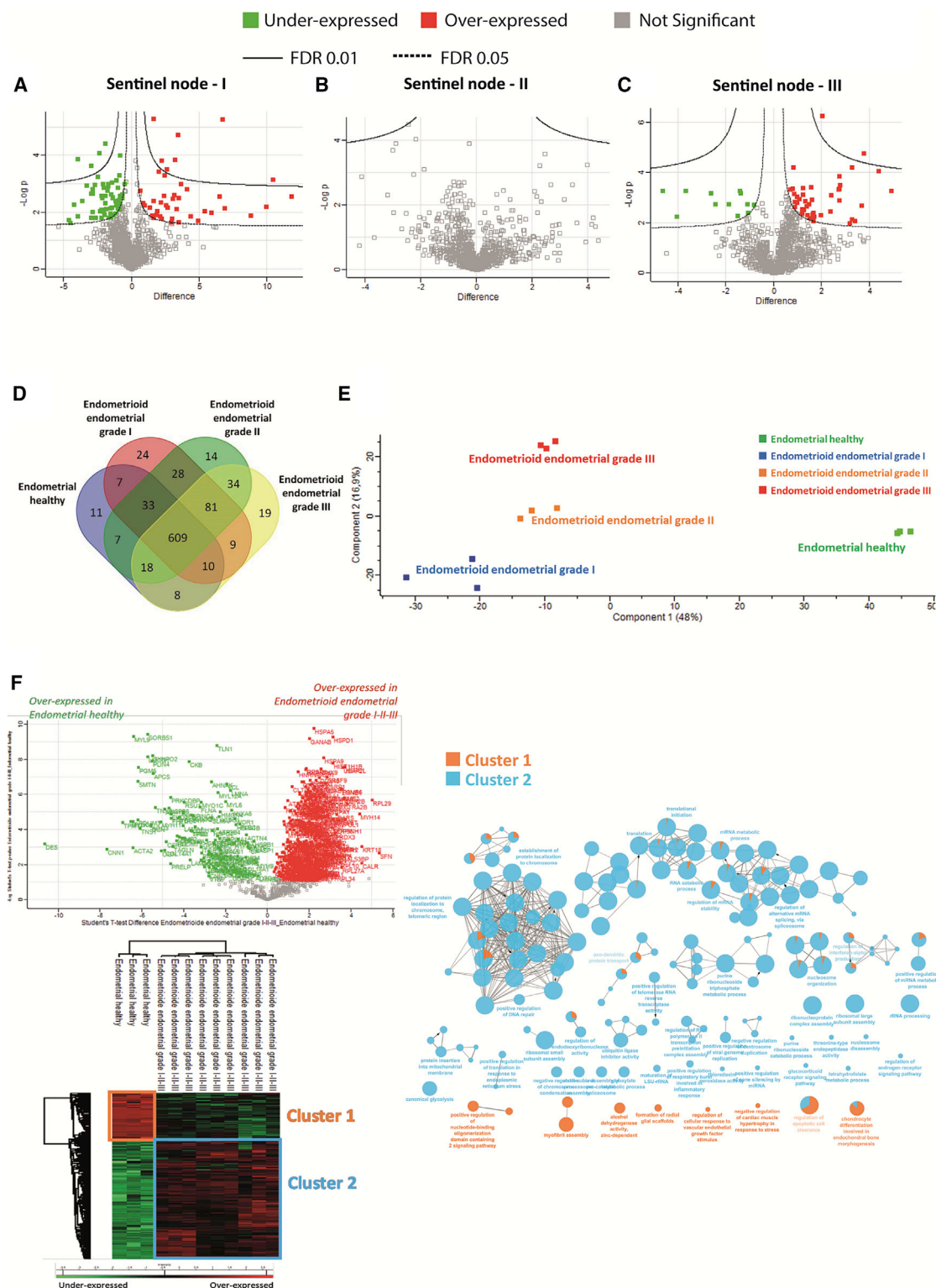
(B) PCA PC1 versus PC3 of the proteomics data obtained from each grade of SN versus healthy tissue.

(C) Volcano plot of overexpressed proteins in normal SNs versus ones in carcinoma (grades I–III).

(D) STRING analysis of proteins present in cluster 1 corresponding to proteins overexpressed only in grade I SNs.

(E) Funrich biological process analyses of normal versus carcinoma SN.

(F) GSEAs associated with Cytoscape of the four clusters.



**Figure 3. Establishment of a molecular transition between SLN and EC grades**

(A–C) Volcano plots of proteins under- or overexpressed in (A) grade I SNs, (B) grade II SNs, and (C) grade III SNs.  
(D) Venn diagram representing specific proteins per grade versus normal and ones in common from SNs.  
(E) PCA PC1 versus PC2 of EC versus normal tissue.

(legend continued on next page)



immune factors involved in the innate immune response. The transition from grade I to grade II can be explained by the switch from an immune profile to cytoskeleton modifications and proliferation. We then refined the analysis with a false discovery rate (FDR) of less than 0.01 to compare grade II with grade III SLNs (Figure S2F). Nineteen proteins were identified to be in common between grade II and grade III; most of these proteins are involved in metabolism and in the Warburg effect (Figure S2F, inset STRING analysis). Cancer cell evolution to epithelial-mesenchymal transition (EMT) is suggested, as reflected by the presence of hypoxia upregulated protein 1. However, resistance of the immune response to cancer cells is still present in grade III. In fact, regulators of self-antigen presentation are still present, such as Tap1, HLA-CW12 (human leukocyte antigen), and class II HLA-DRB1.

### Shotgun proteomics of EC

Proteins from the 12 samples (normal and grades I–III) were extracted and subjected to shotgun proteomics analyses. 1,280 proteins showed a significant difference in expression level. After filtering proteins based on a minimum number of values in at least one of the four defined groups (2/3 of valid values), 913 proteins were obtained (Data S3). According to the Venn diagram, 11 proteins were specific to healthy endometrial tissue, 24 to grade I EC, 14 to grade II, and 19 for grade III (Figure 3D; Table S3). Moreover, principal-component analyses (PCAs) revealed a clear separation between the different grades and healthy endometrial tissue (Figure 3E). The volcano plot associated with the heatmap confirmed the presence of 2 clusters, i.e., a cluster representative of normal endometrial tissue and cluster 2 related to EC (Figure 3F). GSEA associated with Cytoscape confirmed that cluster 2, corresponding to proteins overexpressed in EC, are involved in translation, transcription, and nucleotide metabolism (Figure 3F). Specific proteins identified in grade I EC tissue are related to the immune response with the presence of antimicrobial peptides (neutrophil elastase, neutrophil gelatinase-associated lipocalin, bactericidal permeability-increasing protein, and azurocidin). Among the immune factors, we identified interleukin-16 (IL-16); receptor-type tyrosine-protein phosphatase C, required for T cell activation; Integrin alpha-M; and Integrin ITGAM/ITGB2, known to be implicated in various adhesive interactions of monocytes, macrophages, and granulocytes. We also identified CD74, which is known to play a role in major histocompatibility complex (MHC) class II antigen presentation. However, CD74 is a poor prognostic marker in breast cancer<sup>26</sup> and PTX3 in EC, according to the TCGA (Table S3). For grade II, 4 proteins involved in the Wnt pathway have been identified (i.e., SMARCA4, GNB2, GNB4, and 26S proteasome subunit 10 [PSMC10]) (Table S3) and in grade III, NOTUM and PSMC6. Moreover, endosialin is known to play a role in tumor angiogenesis.<sup>27</sup> The host cell factor c1 (HCFC1) is an immunomodulator that plays a role in limiting the anti-cancer immune response and production of cytokines such as IL-6 or IL-8, which can contribute to neovascularization or tumor growth.<sup>28</sup> Volcano

plots confirmed the high number of specific proteins identified in the 3 stages of EC development (Figures 4A–4C).

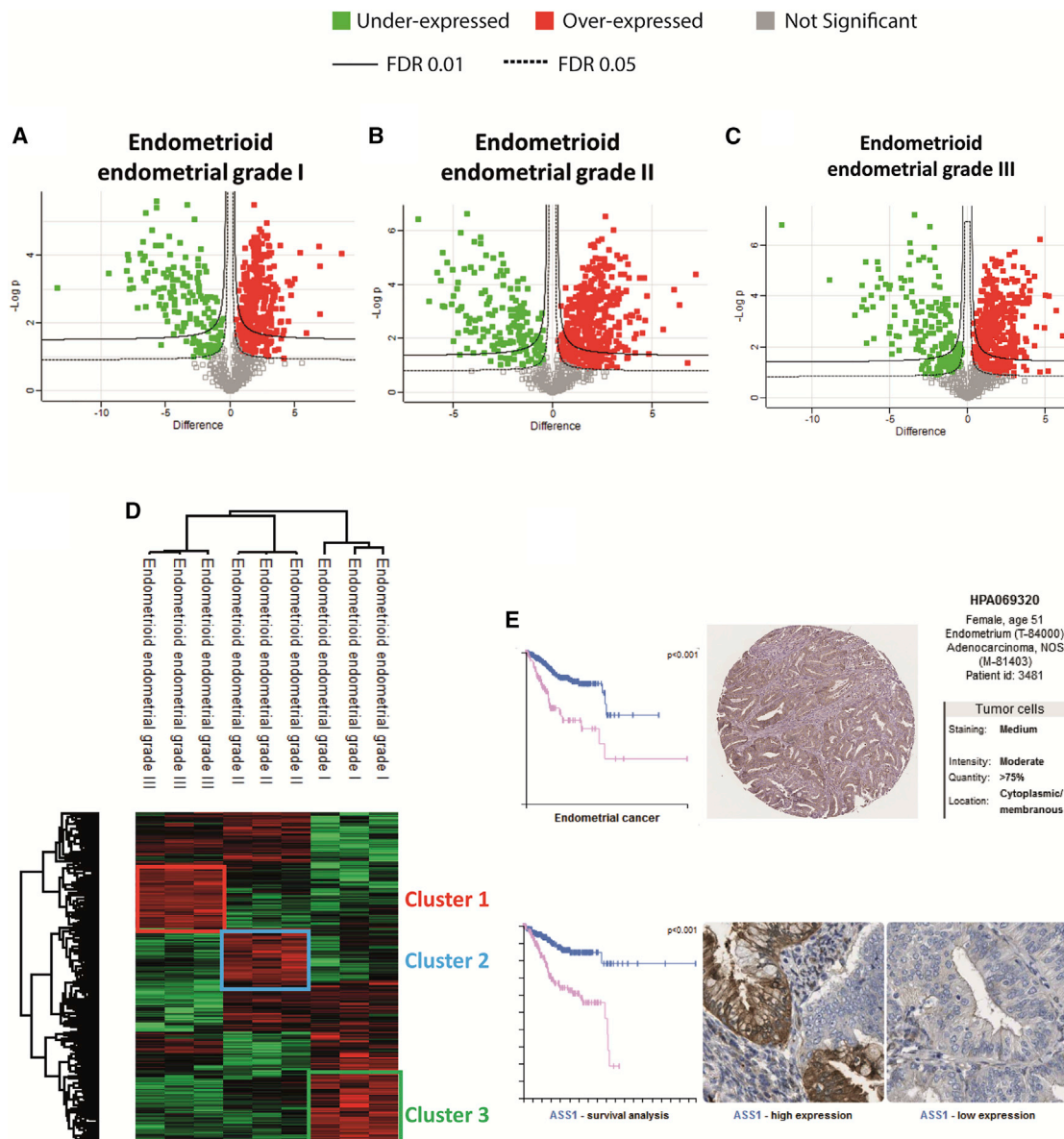
To better understand the modulation registered across the different lesions, a multiple-sample ANOVA with  $p < 0.01$  was performed (Figure 4D). A total of 384 proteins showed a significant difference in expression between the 4 groups. Hierarchical clustering and heatmap representation established a good separation between the three EC grades (Figure 4D; Data S3). Cluster 1 contained proteins overexpressed in grade I tumors. Among the identified proteins, 30 are involved in immune response (Figure S3A), such as gamma interferon-inducible protein 16 (IFI16), lysozyme C, myeloperoxidase, and lactotransferrin. The other ones are related to the metabolism of RNA. For grade II, 42 proteins implicated in cytoskeleton protein binding, the ribosome, and actin-binding were detected in cluster 2 (Figure S3B). A network implicating TPM4, TPM2, TPM1, MYH10, and MYH11 was identified, similar to what we found recently in glioma.<sup>29</sup> For grade III, some of the proteins are involved in cell adhesion (Cadherin-13,  $\beta$ -catenin, tenascin, vitronectin, emilin 1, and collagen alpha-1 chains) and the extracellular matrix, including ApoE and HSPG2 (cluster 3; Figure S3C). ApoE is important for proliferation and survival of ovarian cancer.<sup>30</sup> An analysis with a FDR of 0.01 is presented in Table S4. Interestingly, we identified PTX3 and ASS1 to be overexpressed in grade I and grade II tumors, respectively. These two proteins are considered to be among the top 20 most unfavorable prognostic factors according to TCGA data<sup>17</sup> and antibody-based protein data<sup>18</sup> (Figure 4G).

### Comparison of sentinel node and EC proteomes

A comparison analysis was performed based on the 24 samples used previously after ANOVA with a FDR of 0.01. 659 significant proteins were identified among 1,053. After hierarchical clustering and heatmap representation, healthy endometrial tissue is separated from the sentinel nodes and endometrial tumor grades. Interestingly, one branch regroups sentinel node grade I and endometrial grade I samples. Endometrial grade II and III tumors are also regrouped and separated from sentinel node grades II and III (Table S5). Cluster 1 corresponds to the common proteins overexpressed in sentinel and endometrial grade I tumor tissue (Table S5), and cluster 2 represents the common overexpressed proteins between SLN and EC grade III tissue (Table S5). Cluster 1 contains 22 proteins involved in immune response (Figure 5A) with 10 antimicrobial peptides (neutrophil gelatinase-associated lipocalin, lactotransferrin, lysozyme C, neutrophil elastase, cathepsin G, RNAS3, myeloperoxidase, azurocidin, S100A9, and S100A). Interestingly, cluster 1 also contains the MNDA, which acts as a transcriptional activator/repressor in the myeloid lineage, as well as SerpinB1 and integrins (ITGB2 and ITGA6), also involved in immunity. Besides these immune factors, which are considered favorable prognosis factors for EC, ANXA1 is also overexpressed in grade I SLNs and EC and is considered a favorable prognosis gene. However, ANXA3, ANXA11, LGALS3, FTH1, CP (ceruloplasmin), SERPINB1, FLOT1, and FLOT2 are unfavorable markers for the overall

(F) Volcano plot and Hierarchical clustering of the most variable proteins between normal tissue and grade I–III EC ( $n = 3$  for each category, ANOVA with permutation-based FDR  $< 0.05$ ) and GSEA analyses of normal versus the 3 grades of the EC associated together. The GSEA was performed on the two clusters identified.





**Figure 4. Proteomics analysis of grade I-III EC**

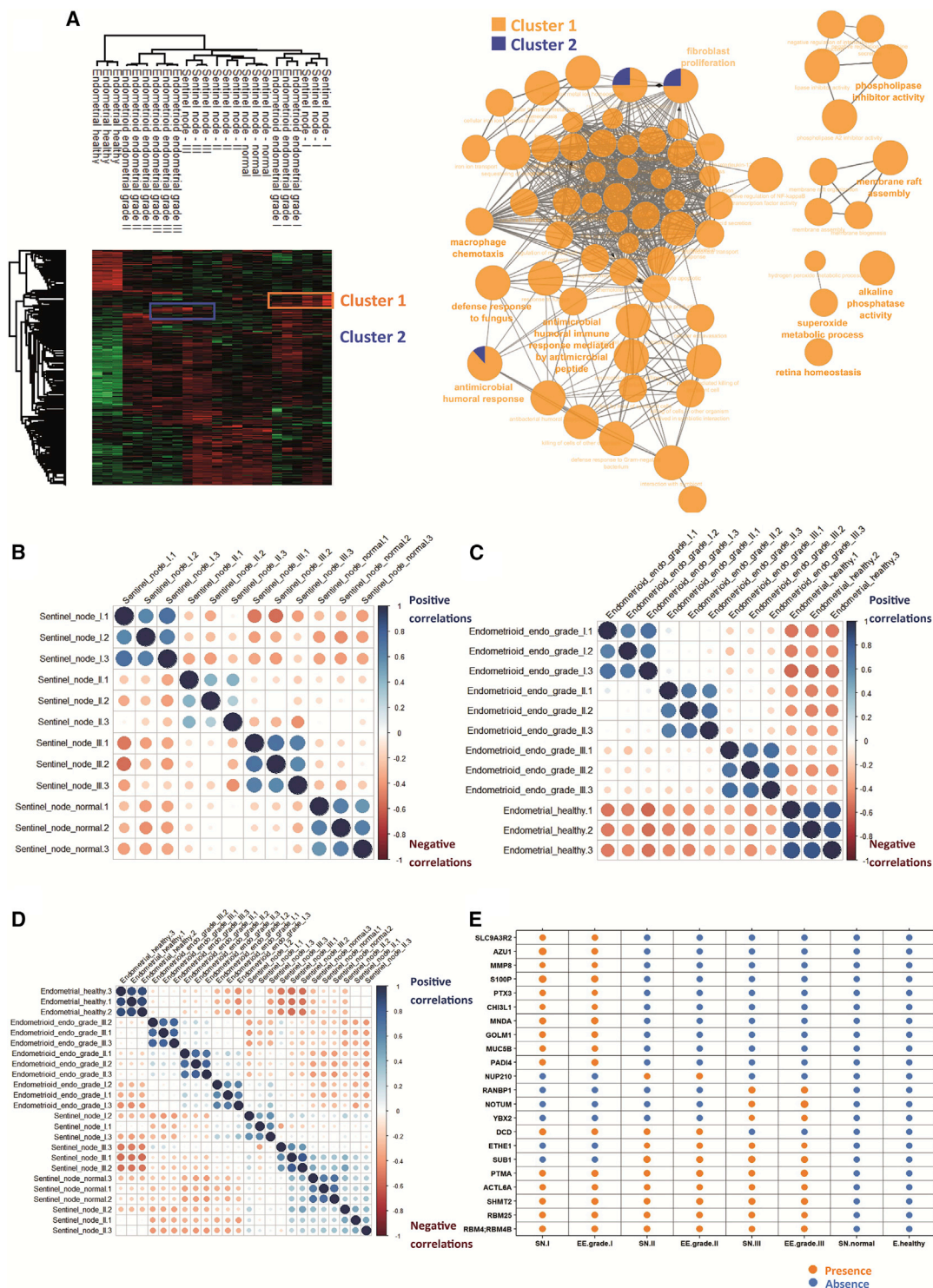
(A–C) Volcano plots of proteins under- and overexpressed in (A) grade I EC, (B) grade II EC, and (C) grade III EC.

(D) Hierarchical clustering of the most variable proteins between grade I-III EC ( $n = 3$  for each category, ANOVA with permutation-based FDR < 0.01).

(E) Survival curve and IHC (immunohistochemistry) of the ASS1 marker obtained from a pathological atlas.

survival (Figure S4A). Cluster 2 contains ALDH2, a favorable prognosis gene for EC, and PRSS3, an unfavorable prognosis marker. YBX1, SLC1A5, ALDOA (aldolase, fructose-bisphosphate A), ATP1A1, and UBA2 are unfavorable OS (overall survival) markers (Figure S4A) and were validated in immunocytochemistry and quantified based on pathological atlas data (Figures S4B–S4B’). Moreover, correlation studies between the different grades of SLNs (Figure 5B) or EC (Figure 5C) and between sentinel nodes and EC grades (Figure 5D) established a positive correlation between grades when the datasets are taken individually. Comparison of sentinel nodes and EC grades points

out a positive correlation between grade I of both tissues. Using correlation matrices (Figure 5E), we found that PTMA (prothymosin alpha), ACTL6A, SHMT2, RBM25, and RBM4 were detected in grade I, II, and III sentinel nodes as well as in grade I, II, and III of EC. SUB1 and ETHE1 are expressed in grade II and III SLNs and grade II and III EC. DCD is present in grade I and II SLNs and grade I and II EC. YBX2, NOTUM, and RANDBP1 are found specifically in grade III SLNs and grade III EC. NUP210 is specific to grade II SLNs and grade II EC. PADI4, MUC5B, GOLM1, MNDA, CHI3L1, PTX3, SP100, MMP8, AZU1, and SLC9A3R2 are specific to grade I SLNs and grade I EC.



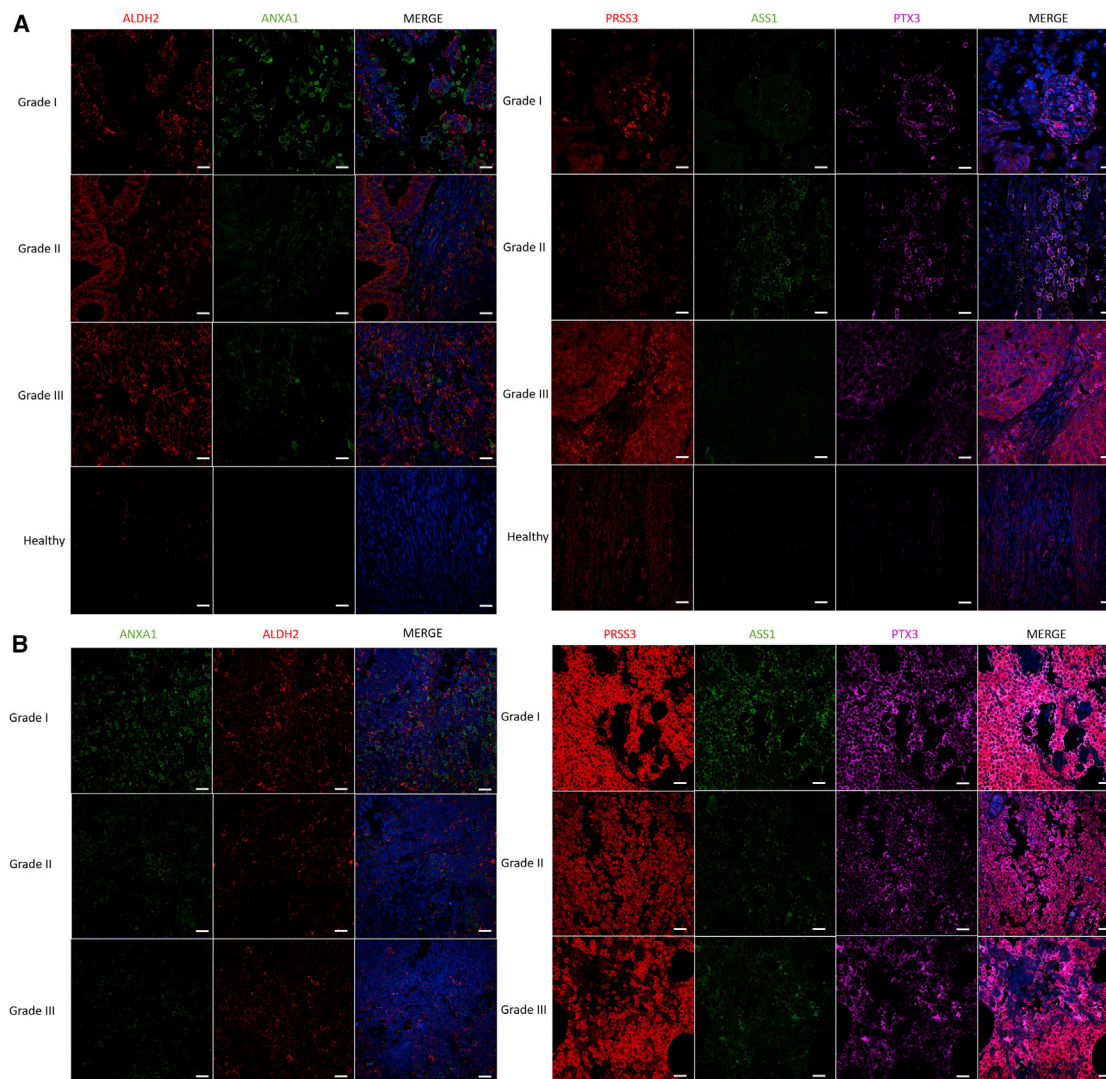
**Figure 5. Similar marker proteins highlighted between SLN and EC according to their grades**

(A) Hierarchical clustering of the most variable proteins between healthy tissue and grade I–III SNs and grade I–III EC ( $n = 3$  for each category), ANOVA with permutation-based FDR  $< 0.01$ . 3 clusters were identified (1–3), and GSEAs are shown as insets for clusters 1 and 2.

(B–D) Matrix (B) correlation studies between the different grades of SNs, and (C) EC grades (D) between SN and EC grades.

(E) Table of selected markers found in TCGA and detected or not detected in SNs and endometrial carcinoma (EC) at different grades as well as normal/healthy tissue.





**Figure 6. IHC validation of the panel of survival markers identified**

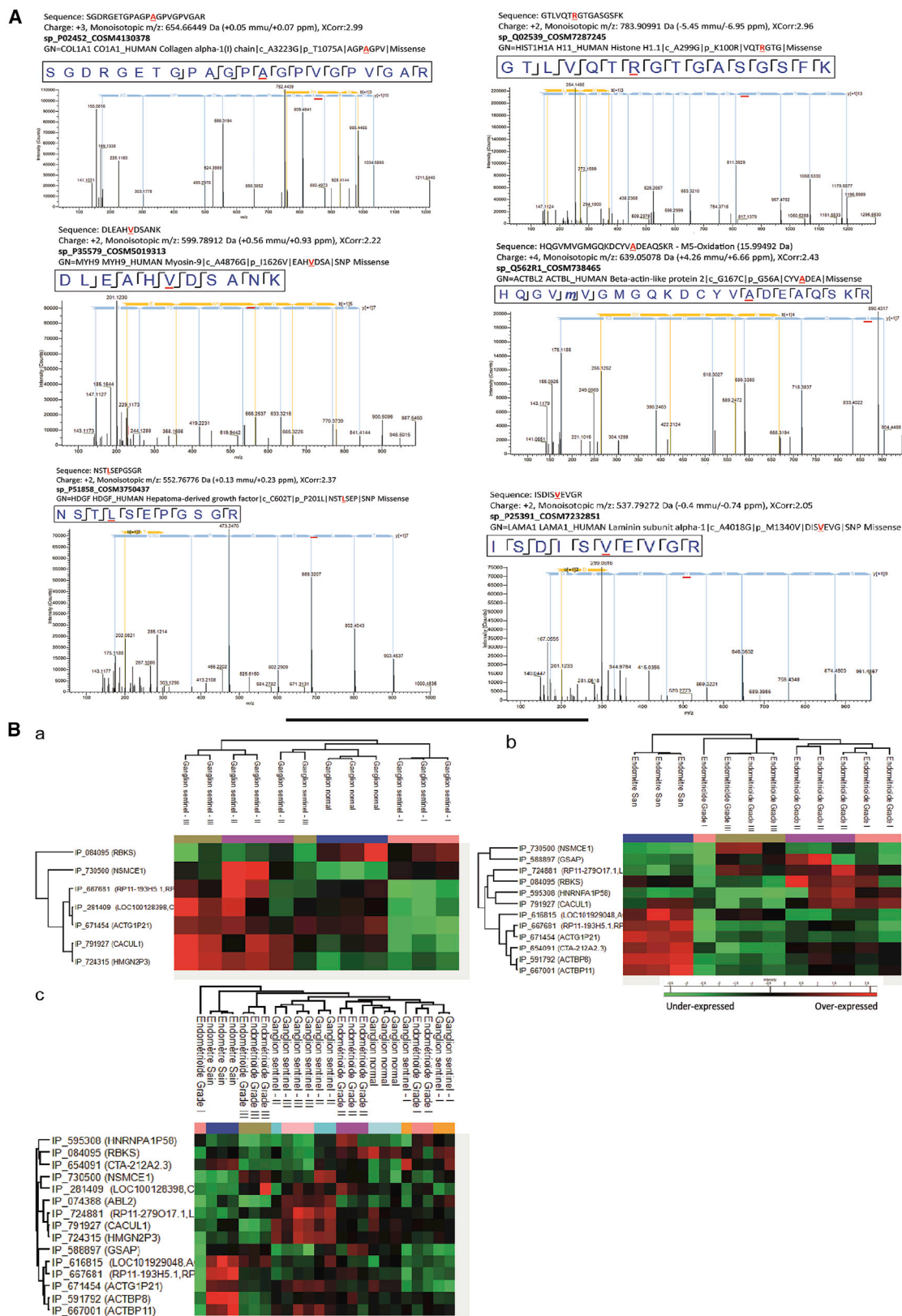
Representative fluorescence images of the 5 proteins of EC grades. PRSS3, PTX3, and ASS1 are associated with poor outcome, and ALDH2 and ANX1 are considered positive outcome markers in (A) EC and (B) SLNs. Images were acquired with a confocal microscope at 40× magnification. Scale bars, 20 μm.

We established the presence of markers common between SLNs and EC tissue that are grade dependent. Among the identified markers, PRSS3, PTX3, and ASS1 are considered poor outcome gene marker, whereas ALDH2 and ANXA1 are positive outcome markers. Validation was performed on 13 samples (Table S1; corresponding to 8 patients) by immunofluorescence (Figure 6). In EC, PRSS3 is more expressed in grade III, ASS1 in grade II, and PTX3 and ANXA1 in grade I, whereas ALDH2 is present in grade II and grade III of EC. In SLNs, PRSS3, PTX3, and ASS1 are highly expressed compared with other markers, such as ANXA1 and ALDH2, which are close to undetectable. ASS1, PTX3, and PRSS3 are highly detectable in grade I, whereas their expression is lower in grades II and III. PRSS3 is slightly less detectable in grade II, and its detection is increased in grade III. For PTX3, the level of detection in grades II and III is lower than in grade I. None of them are detected in healthy pa-

tients (Figures 6A and 6B). The other markers correlated with poor outcome (YBX1, SLC1A5, ALDOA, ATP1A1, and UBA2) have already been tested on transcription-mediated amplification (TMA), and we confirmed their presence at elevated levels (Figures S4B and S4B"). We also detected some protein markers that can be correlated with a positive outcome, associated with longer overall survival; these included the CD74, GOLM1, SLC9A3R2, and SP100 proteins, according to the TCGA.

### Mutations

Several studies have evaluated genetic mutation frequencies in individuals with EC of different tumor grades.<sup>31</sup> In this context, we sought to determine whether these mutations can be translated and detected in proteins. For that purpose, we used a human database combined with the XMan v.2 database.<sup>32</sup> This database contains information about mutated peptides that



(legend on next page)



can be found in some cancers, extracted from the COSMIC database. Sixty-six mutated peptides were identified (Data S4). Among the proteins from which mutated peptides were identified, we retrieved collagen alpha 1, histones, laminin 1, myosin 9, HDGF, CAP-1, fibrinogen beta unit, tenascin, PTBP1, actin, hemoglobin subunits, integrin beta 2, ACTBL, SRP-14, annexin 6, TMP2, HCLS1, and lactotransferrin (Data S4). Laminin subunit alpha-1 mutated peptide (Figure 7A) is only detected in EC grade II tissue, and HDGF protein is identified in SLN grade I (Data S4). Correlation between SLN grades and EC grades has shown that CO1A1, histone H1A, histone 2A2C, myosin 9, ACTBL1, and HDGF are in common with CO1A1 mutated peptides, which were found in all samples (Data S4; Figure 7).

### Ghost proteins

We have previously established the presence of proteins derived from non-canonical human open reading frames promoting cancer metastasis in high-grade serous carcinoma (HGSC) and glioma.<sup>33–37</sup> From our proteomics data, 36 alternative proteins were identified, with more than 80% derived from non-coding RNA (ncRNA), and 10% of the mRNA coding for RefProt is derived from the 5' UTR, 11% from the 3' UTR, and 3% from a shift in the CDS (coding sequence) (Data S5). Eleven are common to all tissues (Data S5). Six are common to all grades of sentinel nodes and grade II of EC tissues (Data S5). Only Alt-RBKS, Alt-LOC100128398, Alt-RP11-193H5.1, Alt-CTA-212A2.3, and Alt-HNRNPA1P58 are not expressed in normal endometrial tissue. Alt-CTA-212A2.3, Alt-HNRNPA1P58, Alt-SLC13A1, and Alt-AP002884.2 are absent in normal lymph nodes (Data S5). A heatmap of the different ghost proteins demonstrates the presence of 7 alternative proteins in sentinel nodes.

Alt-HMGN2P3 and Alt-CACUL1 are overexpressed in grade II and III SLNs. The other ones were not quantified or overexpressed (Figure 7Ba). In endometrial tissue, Alt-ACTBP11, Alt-ACTBP8, Alt-CTA-212A2.3, Alt-ACTG1P211, and Alt-LOC101929048 are overexpressed in healthy endometrial tissue. Alt-CALCUL1 and Alt-RBKS are overexpressed in grade II. Amt-NSMCE1 is overexpressed in grade III (Figure 7Bb). The combination of all tissues gave some specific clusters. Alt-CALCUL1, Alt-HMGN2P3, and Alt-RP11-279O17 are overexpressed only in SLN grades II and III. Alt-ACTBP11, Alt-ACTBP8, Alt-LOC101929048, and Alt-RP11-193H5 are only overexpressed in healthy endometrial tissue. No clear alternative proteins in common in sentinel nodes and endometrial tissue can be found (Figure 7Bc).

### DISCUSSION

In this work, based on spatially resolved proteomics analysis of grade I–III clinical samples derived from individuals with EC and SLNs, we establish the presence of common proteomics markers that are grade dependent. Compared with recent studies using SWATH-MS (Sequential Window Acquisition of

all Theoretical Mass Spectra) proteomics on grade II and grade I EC, we identified all of their validated markers; i.e., CAPS, PKM, AZU1, CNN1, S100A8, STMN1, and CTSG proteins.<sup>38</sup> These proteins are known to have interactions with drugs; some of these are US Food and Drug Administration (FDA)-approved drugs, such as SMARCA4.<sup>39</sup> However, these markers are not used from SLN analyses to predict EC status. In our study, we identify markers that cross-correlate nodal status with EC grade. Among the markers identified in EC and SLNs, 3 markers (PRSS3, PTX3, and ASS1) are considered poor outcome gene markers in EC, and 2 (ALDH2 and ANXA1) are positive outcome markers. These markers are linked to tumor cell motility. In fact, PTX3 contributes to melanoma cell invasion<sup>40</sup> as well as in other cancers,<sup>41</sup> through a Toll-like receptor 4 (TLR4)/nuclear factor  $\kappa$ B (NF- $\kappa$ B) signaling pathway. ASS1 protein is required for cancer cell migration.<sup>42,43</sup> PRSS3 is also known to be involved in tumor metastasis.<sup>44</sup> PRSS3 upregulates VEGF (vascular endothelial growth factor) expression via the PAR1-mediated ERK (extracellular signal-regulated kinases) pathway and promote tumors progression and metastasis.<sup>45</sup> The two positive outcome markers (ANXA1 and ALDH2) are also involved in tumor motility, but in its modulation. ALDH2 is known as one of the key regulators in tumor metastasis, especially in the lungs.<sup>46</sup> ALDH2 functions as a mitogen-activated protein kinase (MAPK) upstream to inhibit cell proliferation and migration, promote cell apoptosis, and alter EMT by elevating E-cadherin and attenuating vimentin. The role of ANXA1 in tumor motility and metastasis is still unclear.<sup>47</sup> Loss of annexin A1 expression has been observed in breast, gastric, esophageal, prostate, bladder, head and neck, laryngeal, and oral cancer and correlates with tumorigenesis and malignant tendency. However, its expression has also been linked to advanced stages of specific cancers as well as metastatic tendency and degree of differentiation. In this case, ANXA1 expression increased inversely to epithelial markers such as E-cadherin and cytokeratin (CK) 8 and 18 and proportionally to mesenchymal ones such as vimentin, ezrin, and moesin.<sup>48</sup> ANXA1 seems to regulate metastasis by favoring cell migration/invasion intracellularly as a cytoskeleton remodeling factor and extracellularly as a ligand of formyl peptide receptor.<sup>49,50</sup> Most of these markers have not been reported previously in EC at the protein level, based on IHC data, except the ASS1 protein. We validate these markers using immunofluorescence and confirmed their specific presence in EC, offering new tools for pathologists for diagnosis. Other markers identified as poor outcome markers (YBX1, SLC1A5, ALDOA, ATP1A1, and UBA2) have already been tested on TMA, and their presence in high abundance was confirmed, as we quantified (Figure S4B). We also detected some positive outcome markers (CD74, GOLM1, SLC9A3R2, and SP100) according to overall survival (from the TCGA and pathological protein atlas).<sup>17,18</sup> These results provide a clear answer regarding the disagreement among cancer centers

### Figure 7. Mutation and alternative proteins analyses

(A) MS/MS spectrum of mutated peptides.

(B) Hierarchical clustering of the most variable alternative proteins between healthy tissue and grade I–III (a;  $n = 3$  for each category, ANOVA with permutation-based FDR < 0.01), healthy tissue and grade I–III EC (b;  $n = 3$  for each category, ANOVA with permutation-based FDR < 0.01), and between healthy tissue and grade I–III SLNs and grade I–III EC ( $n = 3$  for each category, ANOVA with permutation-based FDR < 0.01).

regarding the value of lymph node dissection. It is clear that lymph node metastasis is one of the most important prognostic factors in EC. EC and SLN data identified key proteins that can be used to establish this correlation. In fact, PTMA, ACTL6A, SHMT2, RBM25, and RBM4 are detected in SLN and EC grades I and II and in EC grade III, whereas SUB1 and ETHE1 are expressed in SLN and EC grades II and III.

Of interest, the DCD protein is found in SLN and EC grades I and II. The YBX2, NOTUM, and RANDBP1 proteins are found specifically in SLN and EC grade III. NUP210 is specific to SLN and EC grade II. PADI4, MUC5B, GOLM1, MNDA, CHI3L1, PTX3, SP100, MMP8, AZU1, and SLC9A3R2 are specific to grade I SLNs and EC. We can also include proteins with specific mutations as signature markers. Col1A1, histone H1A, histone 2A2C, myoglobin 9, ACTBL1, and HDGF were detected in all cancer tissue, whatever the grade, but lamin1 mutated peptides are only detected in grade II EC tissue and HDGF only in SLN grade I. In addition to these data, another family of proteins has also been identified (i.e., Alt-CALCUL1, Alt-HMGN2P3, and Alt-RP11-279O17) that is overexpressed only in SLN grades II and III.

Based on these specific markers identified in EC and SLNs and ones that are specific to grade I, II, or III, stratification of affected individuals can be established, which will guide treatment decisions. As we detected the 5 markers which were analyzed blind in SLN tissues and validated in EC tissues, SLN mapping is feasible and can accurately predict nodal status in women with EC based on our markers. In this way, the exact underlying molecular mechanisms relating SLN and EC grades have now been highlighted by systemic biology-proteomics-based study, which facilitates accurate EC detection and can be used as a therapeutic endpoint target.

### Limitations of study

The study was performed on a cohort of 41 samples from 15 individuals. A larger cohort is needed to definitively establish the markers we identified and validated for routine molecular pathology. Such a large validation cohort is under construction at the national level and will be tested as a multicentric assay through a national PHCR (Programme hospitalier de recherche clinique) from the National Institute of Cancer.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - Antigen retrieval
  - Microproteomic analysis
  - Trypsin digestion on tissue
  - Liquid extraction
  - NanoLC-ESI-MS<sup>2</sup>

### ● QUANTIFICATION AND STATISTICAL ANALYSIS

- Data interrogation and analyses
- Subnetwork Enrichment Pathway Analyses and statistical Testing
- Mutation identification
- Ghost proteins identification
- Confirmatory immunohistochemistry analyses

### ● ADDITIONAL RESOURCES

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2021.100318>.

### ACKNOWLEDGMENTS

This research was supported by grants from the Ministère de L'Education Nationale, de L'Enseignement Supérieur et de la Recherche and SIRIC ONCOLille (to I.F.) and INCA-DGOS-INSERM 6041aa, Inserm, and the PHRC SENTIRAD (to E.L.).

### AUTHOR CONTRIBUTIONS

I.F., M.W., and M.S. conceived the study. M.W., P.S., S.A., and M.D. performed the experiments. E.L. and F.N. performed surgery. The pathology analysis was done by A.-S.L. F.K. performed the systemic biology analyses. S.A. and T.C. performed data analyses. I.F. and M.S. supervised the project, participated in experimental design and data analyses, and wrote the manuscript with contributions from all co-authors. M.S. and I.F. acquired funding for the project.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 22, 2020

Revised: December 17, 2020

Accepted: May 20, 2021

Published: June 15, 2021

### REFERENCES

1. Berveiller, P., Mir, O., Veyrie, N., and Barranger, E. (2010). The sentinel-node concept: a dramatic improvement in breast-cancer surgery. *Lancet Oncol.* 11, 906.
2. Gould, E.A., Winship, T., Philbin, P.H., and Kerr, H.H. (1960). Observations on a "sentinel node" in cancer of the parotid. *Cancer* 13, 77–78.
3. Abu-Rustum, N.R. (2014). Sentinel lymph node mapping for endometrial cancer: a modern approach to surgical staging. *J. Natl. Compr. Canc. Netw.* 12, 288–297.
4. Cabanas, R.M. (1977). An approach for the treatment of penile carcinoma. *Cancer* 39, 456–466.
5. Cascinelli, N., Belli, F., Santinami, M., Fait, V., Testori, A., Ruka, W., Cavaliere, R., Mozzillo, N., Rossi, C.R., MacKie, R.M., et al. (2000). Sentinel lymph node biopsy in cutaneous melanoma: the WHO Melanoma Program experience. *Ann. Surg. Oncol.* 7, 469–474.
6. Tanis, P.J., Nieweg, O.E., Valdés Olmos, R.A., Th Rutgers, E.J., and Kroon, B.B. (2001). History of sentinel node and validation of the technique. *Breast Cancer Res.* 3, 109–112.
7. Mais, V., Cirronis, M.G., Piras, B., Silvetti, E., Cossu, E., and Melis, G.B. (2011). Intraoperative lymphatic mapping techniques for endometrial cancer. *Expert Rev. Anticancer Ther.* 11, 83–93.
8. Burke, T.W., Levenback, C., Tornos, C., Morris, M., Wharton, J.T., and Gershenson, D.M. (1996). Intraabdominal lymphatic mapping to direct

- selective pelvic and paraaortic lymphadenectomy in women with high-risk endometrial cancer: results of a pilot study. *Gynecol. Oncol.* 62, 169–173.
9. Abu-Rustum, N.R. (2013). The increasing credibility of sentinel lymph node mapping in endometrial cancer (Springer).
  10. Touboul, C., Bentivegna, E., Uzan, C., Gouy, S., Pautier, P., Lhomme, C., Duvillard, P., Haie-Meder, C., and Morice, P. (2013). Sentinel lymph node in endometrial cancer: a review. *Curr. Oncol. Rep.* 15, 559–565.
  11. Smith, A.J.B., Fader, A.N., and Tanner, E.J. (2017). Sentinel lymph node assessment in endometrial cancer: a systematic review and meta-analysis. *Am. J. Obstet. Gynecol.* 216, 459–476.e10.
  12. Rossi, E.C., Kowalski, L.D., Scalici, J., Cantrell, L., Schuler, K., Hanna, R.K., Method, M., Ade, M., Ivanova, A., and Boggess, J.F. (2017). A comparison of sentinel lymph node biopsy to lymphadenectomy for endometrial cancer staging (FIRES trial): a multicentre, prospective, cohort study. *Lancet Oncol.* 18, 384–392.
  13. Liu, S.-L., Han, Y., Zhang, Y., Xie, C.-Y., Wang, E.-H., Miao, Y., Li, H.-Y., Xu, H.-T., and Dai, S.-D. (2012). Expression of metastasis-associated protein 2 (MTA2) might predict proliferation in non-small cell lung cancer. *Target. Oncol.* 7, 135–143.
  14. Park, J.-O., Jung, C.-K., Sun, D.-I., Joo, Y.-H., and Kim, M.-S. (2011). Relationships between metastasis-associated protein (MTA) 1 and lymphatic metastasis in tonsil cancer. *Eur. Arch. Otorhinolaryngol.* 268, 1329–1334.
  15. Jang, K.-S., Paik, S.S., Chung, H., Oh, Y.-H., and Kong, G. (2006). MTA1 overexpression correlates significantly with tumor grade and angiogenesis in human breast cancers. *Cancer Sci.* 97, 374–379.
  16. Xu, X., Kong, X., Liu, T., Zhou, L., Wu, J., Fu, J., Wang, Y., Zhu, M., Yao, S., Ding, Y., et al. (2019). Metastasis-associated protein 1, modulated by miR-30c, promotes endometrial cancer progression through AKT/mTOR/4E-BP1 pathway. *Gynecol. Oncol.* 154, 207–217.
  17. Cancer Genome Atlas Research Network; Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120.
  18. Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfante, R., Arif, M., Liu, Z., Edfors, F., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357, eaan2507.
  19. Sica, A., Porta, C., Amadori, A., and Pastò, A. (2017). Tumor-associated myeloid cells as guiding forces of cancer cell stemness. *Cancer Immunol. Immunother.* 66, 1025–1036.
  20. Shen, W., Xie, J., Zhao, S., Du, R., Luo, X., He, H., Jiang, S., Hao, N., Chen, C., Guo, C., et al. (2018). ICAM3 mediates inflammatory signaling to promote cancer cell stemness. *Cancer Lett.* 422, 29–43.
  21. Werner, H.M.J., and Salvesen, H.B. (2014). Current status of molecular biomarkers in endometrial cancer. *Curr. Oncol. Rep.* 16, 403.
  22. Reyes, H.D., Bosquet, J.G., Stephan, J.M., McDonald, M.E., and Leslie, K.K. (2015). Stathmin over-expression correlates with poor prognosis in patients with endometrial cancer. *Gynecologic Oncology* 137, 66.
  23. Liu, L., Jiao, J., Wang, Y., Zhang, D., Wu, J., and Huang, D. (2014). Lack of association of the TP53BP1 Glu353Asp polymorphism with risk of cancer: a systematic review and meta-analysis. *PLoS One* 9, e90931.
  24. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
  25. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431–432.
  26. Wang, Z.-Q., Milne, K., Webb, J.R., and Watson, P.H. (2017). CD74 and intratumoral immune response in breast cancer. *Oncotarget* 8, 12664–12674.
  27. Rettig, W.J., Garin-Chesa, P., Healey, J.H., Su, S.L., Jaffe, E.A., and Old, L.J. (1992). Identification of endosialin, a cell surface glycoprotein of vascular endothelial cells in human cancer. *Proc. Natl. Acad. Sci. USA* 89, 10832–10836.
  28. Lahad, J.P., Mills, G.B., and Coombes, K.R. (2005). Stem cell-ness: a “magic marker” for cancer. *J. Clin. Invest.* 115, 1463–1467.
  29. Cardon, T., Franck, J., Coyaude, E., Laurent, E.M.N., Damato, M., Maffia, M., Vergara, D., Fournier, I., and Salzet, M. (2020). Alternative proteins are functional regulators in cell reprogramming by PKA activation. *Nucleic Acids Res.* 48, 7864–7882.
  30. Chen, Y.-C., Pohl, G., Wang, T.-L., Morin, P.J., Risberg, B., Kristensen, G.B., Yu, A., Davidson, B., and Shih, I.-M. (2005). Apolipoprotein E is required for cell proliferation and survival in ovarian cancer. *Cancer Res.* 65, 331–337.
  31. Althubiti, M.A. (2019). Mutation frequencies in endometrial Cancer patients of different ethnicities and tumor grades: an analytical study. *Saudi J. Med. Med. Sci.* 7, 16–21.
  32. Aguilera Flores, M., and Lazar, I.M. (2020). XMAN v2—a database of Homo sapiens mutated peptides. *Bioinformatics* 36, 1311–1313.
  33. Vergara, D., Verri, T., Damato, M., Trerotola, M., Simeone, P., Franck, J., Fournier, I., Salzet, M., and Maffia, M. (2020). A Hidden Human Proteome Signature Characterizes the Epithelial Mesenchymal Transition Program. *Curr. Pharm. Des.* 26, 372–375.
  34. Cardon, T., Hervé, F., Delcourt, V., Roucou, X., Salzet, M., Franck, J., and Fournier, I. (2020). Optimized Sample Preparation Workflow for Improved Identification of Ghost Proteins. *Anal. Chem.* 92, 1122–1129.
  35. Cardon, T., Salzet, M., Franck, J., and Fournier, I. (2019). Nuclei of HeLa cells interactomes unravel a network of ghost proteins involved in proteins translation. *Biochim. Biophys. Acta, Gen. Subj.* 1863, 1458–1470.
  36. Delcourt, V., Brunelle, M., Roy, A.V., Jacques, J.F., Salzet, M., Fournier, I., and Roucou, X. (2018). The protein coded by a short open reading frame, not by the annotated coding sequence is the main gene product of the dual-coding gene MIEF1. *Mol. Cell. Proteomics* 17, 2402–2411.
  37. Vanderperre, B., Lucier, J.F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F.M., and Roucou, X. (2013). Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS ONE* 8, e70698.
  38. Janacova, L., Faktor, J., Capkova, L., Paralova, V., Pospisilova, A., Podhorec, J., Ebhardt, H.A., Hrstka, R., Nenutil, R., Aebbersold, R., and Bouchal, P. (2020). SWATH-MS Analysis of FFPE Tissues Identifies Stathmin as a Potential Marker of Endometrial Cancer in Patients Exposed to Tamoxifen. *J. Proteome Res.* 19, 2617–2630.
  39. Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C., et al.; Clinical Proteomic Tumor Analysis Consortium (2020). Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* 183, 1436–1456.e31.
  40. Rathore, M., Girard, C., Ohanna, M., Tichet, M., Ben Jouira, R., Garcia, E., Larbret, F., Gesson, M., Audebert, S., Lacour, J.-P., et al. (2019). Cancer cell-derived long pentraxin 3 (PTX3) promotes melanoma migration through a toll-like receptor 4 (TLR4)/NF- $\kappa$ B signaling pathway. *Oncogene* 38, 5873–5889.
  41. Doni, A., Stravalaci, M., Inforzato, A., Magrini, E., Mantovani, A., Garlanda, C., and Bottazzi, B. (2019). The long pentraxin PTX3 as a link between innate immunity, tissue remodeling, and cancer. *Front. Immunol.* 10, 712.
  42. Tsai, C.Y., Chi, H.C., Chi, L.M., Yang, H.Y., Tsai, M.M., Lee, K.F., Huang, H.W., Chou, L.F., Cheng, A.J., Yang, C.W., et al. (2018). Argininosuccinate synthetase 1 contributes to gastric cancer invasion and progression by modulating autophagy. *FASEB J.* 32, 2601–2614.
  43. Shan, Y.-S., Hsu, H.-P., Lai, M.-D., Yen, M.-C., Chen, W.-C., Fang, J.-H., Weng, T.-Y., and Chen, Y.-L. (2015). Argininosuccinate synthetase 1 suppression and arginine restriction inhibit cell migration in gastric cancer cell lines. *Sci. Rep.* 5, 9783.
  44. Ma, H., Hockla, A., Mehner, C., Coban, M., Papo, N., Radisky, D.C., and Radisky, E.S. (2019). PRSS3/Mesotrypsin and kallikrein-related peptidase

- 5 are associated with poor prognosis and contribute to tumor cell invasion and growth in lung adenocarcinoma. *Sci. Rep.* **9**, 1844.
45. Jiang, G., Cao, F., Ren, G., Gao, D., Bhakta, V., Zhang, Y., Cao, H., Dong, Z., Zang, W., Zhang, S., et al. (2010). PRSS3 promotes tumour growth and metastasis of human pancreatic cancer. *Gut* **59**, 1535–1544.
46. Yang, M., Wang, A., Li, C., Sun, J., Yi, G., Cheng, H., Liu, X., Wang, Z., Zhou, Y., Yao, G., et al. (2020). Methylation-induced silencing of ALDH2 facilitates lung adenocarcinoma bone metastasis by activating the MAPK pathway. *Front. Oncol.* **10**, 1141.
47. Biaoxue, R., Xiguang, C., and Shuanying, Y. (2014). Annexin A1 in malignant tumors: current opinions and controversies. *Int. J. Biol. Markers* **29**, e8–e20.
48. Yang, W., Wang, K., Ma, J., Hui, K., Lv, W., Ma, Z., Huan, M., Luo, L., Wang, X., and Li, L. (2021). Inhibition of Androgen Receptor Signaling Promotes Prostate Cancer Cell Migration via Upregulation of Annexin A1 Expression. *Arch. Med. Res.* **52**, 174–181.
49. Pessolano, E., Belvedere, R., Bizzarro, V., Franco, P., Marco, I., Petrella, F., Porta, A., Tosco, A., Parente, L., Perretti, M., and Petrella, A. (2019). Annexin A1 contained in extracellular vesicles promotes the activation of keratinocytes by Mesoglycan effects: an autocrine loop through FPRs. *Cells* **8**, 753.
50. Belvedere, R., Bizzarro, V., Popolo, A., Dal Piaz, F., Vasaturo, M., Picardi, P., Parente, L., and Petrella, A. (2014). Role of intracellular and extracellular annexin A1 in migration and invasion of human pancreatic carcinoma cells. *BMC Cancer* **14**, 961.
51. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372.
52. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740.
53. Pathan, M., Keerthikumar, S., Ang, C.S., Gangoda, L., Quek, C.Y., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., et al. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **15**, 2597–2601.
54. Vizcaino, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dienes, J.A., Sun, Z., Farrah, T., Bandeira, N., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226.
55. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805.
56. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526.
57. Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguéz, P., Doerks, T., Stark, M., Muller, J., Bork, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568.
58. Bonnet, A., Lagarrigue, S., Liaubet, L., Robert-Granié, C., Sancristobal, M., and Tosser-Klopp, G. (2009). Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares. *BMC Proc.* **3** (Suppl 4), S11.
59. Yuryev, A., Kotelnikova, E., and Daraselia, N. (2009). Ariadne's ChemEffect and Pathway Studio knowledge base. *Expert Opin. Drug Discov.* **4**, 1307–1318.
60. Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Goat polyclonal anti-ASS1	Abcam	Cat# ab77590, RRID: AB_2060467
Rabbit polyclonal anti-PRSS3	Thermo Fisher Scientific	Cat# PA5-23991, RRID: AB_2541491
Rat monoclonal anti-PTX3	Abcam	Cat# ab90806, RRID: AB_2050229
Rabbit polyclonal anti-ALDH2	Thermo Fisher Scientific	Cat# PA5-27414, RRID: AB_2544890
Mouse monoclonal anti-ANXA1	OriGene	Cat# TA500967, RRID: AB_11125833
<b>Biological samples</b>		
Endometrioid biopsies from human patient	Centre Oscar Lambret (Lille France)	Sentirad-1502, EudraCT: 2015-001732-38
Sentinel nodes biopsies from human patient	Centre Oscar Lambret (Lille France)	Sentirad-1502, EudraCT: 2015-001732-38
<b>Chemicals, peptides, and recombinant proteins</b>		
Ammonium bicarbonate (NH <sub>4</sub> HCO <sub>3</sub> )	Sigma-Aldrich	09830; CAS: 1066-33-7
HPLC water	Biosolve	232106; CAS: 7732-18-5
Formic acid (FA)	Sigma-Aldrich	695076; CAS: 64-18-6
Trifluoroacetic acid (TFA)	Sigma-Aldrich	T62200; CAS: 76-05-1
Methanol (MeOH)	VWR	85800; CAS: 67-56-1
Xylene	Sigma-Aldrich	Cat#247642
Acetonitrile (ACN)	VWR	20060.320 CAS: 75-05-8
Porcine modified trypsin	Promega	Cat#V5111
Tris	Interchim	UP031657; CAS: 77-86-1
<b>Deposited data</b>		
Raw and analyzed data	This paper	ProteomeXchange: PXD020410
<b>Software and algorithms</b>		
MaxQuant v1.5.6.5	Cox and Mann <sup>51</sup>	<a href="https://www.maxquant.org/">https://www.maxquant.org/</a>
Perseus v1.6.12.0	Tyanova et al. <sup>52</sup>	<a href="http://maxquant.net/perseus/">http://maxquant.net/perseus/</a>
Proteome Discoverer 2.3	Thermo Scientific	<a href="https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/teome-discoverer-software.html">https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/teome-discoverer-software.html</a>
Python 3.8.3	Python Software Foundation (PSF)	<a href="https://www.python.org/downloads/">https://www.python.org/downloads/</a>
R-4.0.0	R Core Team	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
Cytoscape 3.6.0	Cytoscape Consortium	<a href="https://cytoscape.org/">https://cytoscape.org/</a>
FunRich	Pathan et al. <sup>53</sup>	<a href="http://www.funrich.org/">http://www.funrich.org/</a>
ImageJ	NIH	<a href="https://imagej.nih.gov/ij/download.html">https://imagej.nih.gov/ij/download.html</a>
<b>Other</b>		
Ghost protein database (GRCh38.83, GRCh38.p7)	Openprot	<a href="https://openprot.org/">https://openprot.org/</a>
XManV2 database	Aguilera Flores and Lazar <sup>32</sup>	<a href="https://github.com/lazarlab/XMan-v2">https://github.com/lazarlab/XMan-v2</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Michel Salzet ([michel.salzet@univ-lille.fr](mailto:michel.salzet@univ-lille.fr)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

The raw data and result files used for analysis were deposited at the ProteomeXchange Consortium<sup>54</sup> (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD020410.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

A cohort of 41 samples was selected for this study including 6 healthy Endometrium, 4 Grade I endometrioid, 8 Grade II endometrioid, 4 Grade III endometrioid, 3 Normal sentinel nodes, 4 Grade I sentinel nodes, 8 Grade II sentinel nodes, 4 Grade III sentinel nodes. The age, type of tumors and patient information has been described in [Table S1](#). The normal tissue was analyzed from a healthy normal tissue section (presenting no abnormalities on IHC) from a patient with p53 signature lesion. Prior to the experiments, patients were asked to sign an informed consent, authorization form describing the experimental protocol and instrument and exposure to the hazards. No personal information and data, such as the name of the individuals and identifiers were used in these experiments. A randomized number was assigned to everyone. This clinical study (Sentirad-1502, EudraCT: 2015-001732-38): Randomized study comparing sentinel node (SN) policy to current French initial staging protocols in early-stage endometrial carcinomas at intermediate and high risk of recurrence is supported by the French cancer clinical research projects funding program 2014 (National Cancer Institute, INCa). The endometrioid and sentinel nodes biopsies were obtained from patients of the Centre Oscar Lambret (Lille, France). All experiments were approved by the local Ethics Committee (CPP Nord-Ouest I on July 20th, 2015, CPP 03/008/2015) in accordance with the French and European legislation on this topic. The study complies with the MR004 reference methodology adopted by the French Data Protection Authority (Paris, France), and we checked that patients did not object to the use of their data and biological samples for research purposes. Serial sections of sentinel nodes were realized every 2–3 mm and then formalin-fixed and paraffin-embedded. On these tissues, HPS and immunostaining of pancytokeratins (CKAE1/3, CKAE1/4, CKAE1/5), P53, L1CAM were performed. Twenty-four samples were selected for spatially resolved shotgun microproteomic analyses ([Table S1](#)). This sample selection has been performed by taking into account the ability to find in the same patient the same Grade in endometrial carcinoma and in sentinel nodes tissues. This cohort is considered a diverse representative of the population as each sample is biologically investigated in triplicate.

## METHOD DETAILS

### Antigen retrieval

All the slides were unmounted, and the resin was removed by soaking them overnight in xylene and rinsing them with xylene and ethanol baths. The tissues are rehydrated using 5' each successive bath of decreasing ethanol degree (2x95°, 1x30°) and two baths of 10mM NH<sub>4</sub>HCO<sub>3</sub> buffer. Then, antigen retrieval was performed to relax the tissue and increase the trypsin access to biomolecules. Slides are dipped in 90°C pH9 20mM Tris for 30 minutes, rinsed in two baths of 10mM NH<sub>4</sub>HCO<sub>3</sub> for 2 minutes each, and dried under vacuum at room temperature.

### Microproteomic analysis

Areas of interest have been selected on the tissue. These areas were digested, extracted and then analyzed in nanoLC-MS.

### Trypsin digestion on tissue

Tryptic digestion was performed using a Chemical Inkjet Printer (CHIP-100, Shimadzu, Kyoto, Japan). The regions were selected from the tissue scanned on the software, and then the trypsin solution (40μg/mL, 50mM NH<sub>4</sub>HCO<sub>3</sub> buffer) was deposited on these region defined to 1mm<sup>2</sup> for 2h. During this time, the trypsin was changed every half-hour to avoid the autolytic digestion. With 350 cycles and 450pl per spot, a total of 6.3μg was deposited. To stop the digestion, TFA 0.1% was spotted during 25 cycles.

### Liquid extraction

After microdigestion, the content of the spot was collected by liquid microjunction using the TriVersa Nanomate (Advion Biosciences Inc., Ithaca, NY, USA) using the parameters of liquid extraction and surface analysis (LESA). 3 Mixtures of different extraction solvents have been prepared and are composed of 0.1% TFA, ACN / 0.1% TFA (8: 2, v / v) and MeOH / 0.1% TFA (7: 3, v / v). A complete LESA sequence run 2 cycles for each mixture solvent. The first step was to aspirate 2μl of solvent into a tip, 0.6 μl was deposited on the tissue to create a liquid microjunction with 10 aspirate-dispense cycles to perform the extraction, and the extracted solution was collected in 0.2 mL weak binding tubes. For each interesting spot, 2 sequences are grouped together in the same bottle.

### NanoLC-ESI-MS<sup>2</sup>

After liquid extraction, samples were freeze-dried in a SpeedVac concentrator (SPD131DPA, ThermoScientific, Waltham, Massachusetts, USA), reconstituted with 10μL 0.1% TFA and subjected to solid-phase extraction to remove salts and concentrate the peptides.

This was done using a C-18 ZipTip protocol (Millipore, Saint-Quentin-en-Yvelines, France). The pipettor was set to 10  $\mu$ L and the ZipTip pipette tip was washing by performing 5 aspirate-dispense cycles in ACN, and equilibrated by 5 aspirate-dispense in TFA 0.1%. To bind peptides, 20 aspirate-dispense was performed in the sample and 10 times in TFA 0.1% to remove salts. Peptides were eluted with 20  $\mu$ L of ACN/0.1% TFA (8:2, v/v) by realizing 20 aspirate-dispense, and then the samples were dried for storage. Before analysis, samples were suspended in 20  $\mu$ L ACN/0.1% FA (2:98, v/v), deposited in nanoLC vials and 10  $\mu$ L were injected for analysis. The separation prior to the MS used online reversed-phase chromatography coupled with a Proxeon Easy-nLC-1000 system (Thermo Scientific) equipped with an Acclaim PepMap trap column (75  $\mu$ m ID x 2 cm, Thermo Scientific) and C18 packed tip Acclaim PepMap RSLC column (75  $\mu$ m ID x 50 cm, Thermo Scientific). Peptides were separated using an increasing amount of acetonitrile (5%–40% over 145 minutes) and a flow rate of 300 nL/min. The separation column was kept at 50°C. The LC eluent was electrosprayed directly from the analytical column and a voltage of 2 kV was applied via the liquid junction of the nanospray source. The chromatography system was coupled to a Thermo Scientific Q-Exactive Orbitrap mass spectrometer. The mass spectrometer was programmed to acquire in a data-dependent mode defined to analyze the 10 most intense ions of MS analysis (Top 10). The survey scans were acquired in the Orbitrap mass analyzer operated at 70,000 (FWHM) resolving power. The MS analysis was performed with an m/z mass range between 300 to 1600, an AGC of 3e6 ions and a maximum injection time of 120 ms. The MS/MS analysis was performed with an m/z mass range between 200 to 2000, an AGC of 50000 ions, a maximum injection time of 60 ms and the resolution was set at 17,500 FWHM. Higher Energy Collision Dissociation (HCD) was set to 30%. Precursors ions with charges states > +1 and < +8 were kept for the fragmentation, with a dynamic exclusion time of 20 s.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Data interrogation and analyses

All MS data were processed with MaxQuant<sup>51,52</sup> (Version 1.5.6.5) using the Andromeda<sup>55</sup> search engine. The proteins were identified by searching MS and MS/MS data against the Decoy version of the complete proteome for *Homo sapiens* in the UniProt database (Release March 2017, 70941 entries) combined with 262 commonly detected contaminants. Trypsin specificity was used for digestion mode, with N-terminal acetylation and methionine oxidation selected as a variable modification. We allowed up to two missed cleavages. Initial mass accuracy of 6 ppm was selected for MS spectra, and the MS/MS tolerance was set to 20 ppm for the HCD data. For the identification parameters, FDR at the peptide spectrum matches (PSM) and protein level was set to 1%, and a minimum of 2 peptides per protein in which 1 was unique. Relative label-free quantification of the proteins was conducted into Max-Quant using the MaxLFQ algorithm<sup>56</sup> with default parameters. Analysis of the identified proteins was performed using Perseus software (<http://maxquant.net/perseus/>) (version 1.6.12.0). The file containing the information from the identification (proteinGroup.txt) was used. Briefly, the LFQ intensity of each sample were downloaded in Perseus and the data matrix was filtered by removing the potential contaminants, reverse and only identified by site. The LFQ intensity was logarithmized ( $\log_2[x]$ ). Categorical annotation of the rows was used to define the different group. Venn diagram and principal component analysis (PCA) were done to compare the protein content of each sample. Statistical multiple-sample tests were performed using ANOVA with a p value of 1%. Normalization was achieved using a Z-score with matrix access by rows. Only proteins that were significant by ANOVA were used. The hierarchical clustering and profile plot of only the statistically significant proteins were all performed and visualized by Perseus. Each protein cluster were selected and extracted for biological analysis. Functional annotation and characterization of the identified proteins were performed using FunRich software (version 3) and STRING (version 9.1, <http://stringdb.org>).<sup>57</sup> Pearson's correlation coefficient and matrix representation were generated in R software using corplot package. Gene Set Enrichment Analysis (GSEA) and Cytoscape software (version 3.6.1) were used for the biological process analysis of the clusters selected from the heatmap.

### Subnetwork Enrichment Pathway Analyses and statistical Testing

The Elsevier's Pathway Studio version 10.0 (Ariadne Genomics/Elsevier) was used to deduce relationships among differentially expressed proteomics protein candidates using the Ariadne ResNet database.<sup>58,59</sup> "Subnetwork Enrichment Analysis" (SNEA) algorithm was selected to extract statistically significant altered biological and functional pathways pertaining to each identified set of protein hits among the different groups. SNEA utilizes Fisher's statistical test set to determine if there are nonrandom associations between two categorical variables organized by specific relationships. Integrated Venn diagram analysis was performed using "the InteractiVenn": a web-based tool for the analysis of complex datasets.<sup>60</sup>

### Mutation identification

Proteins identification was also performed using the mutation-specific database.<sup>32</sup> XMan v2 database contains 2 539 031 mutated peptide sequences from 17 599 *Homo sapiens* proteins (2 377 103 are missense and 161 928 are nonsense mutations). The interrogation was performed by Proteome Discoverer 2.3 software and Sequest HT package, using an iterative method. The precursor mass tolerance was set to 15 ppm and the fragment mass tolerance was set to 0.02 Da. For high confidence result, the false discovery rate (FDR) values were specified to 1%. A filter with a minimum Xcorr of 2 was applied. The generated result file was filtered using a Python script to remove unmutated peptides. All mutations were then manually checked based on MSMS spectra profile.

### Ghost proteins identification

RAW data obtained by nanoLC-MS/MS analysis were analyzed using Proteome Discoverer V2.2 (Thermo Scientific) with the following parameters: Trypsin as an enzyme, 2 missed cleavages, methionine oxidation as a variable modification and carbamidomethylation of cysteines as static modification, Precursor Mass Tolerance: 10 ppm and Fragment mass tolerance: 0.6 Da. The validation was performed using Percolator with an FDR set to 1%. A consensus workflow was then applied for the statistical arrangement, using the high confidence protein identification. The protein database was uploaded from Openprot (<https://openprot.org/>) and included RefProt, novel isoforms, and AltProts predicted from both Ensembl and RefSeq annotations (GRCh38.83, GRCh38.p7).<sup>35</sup>

### Confirmatory immunohistochemistry analyses

Grading group validation was performed using antibodies directed against ALDH2, ANXA1, PRSS3, ASS1 and PTX3. After dewaxing and antigen retrieval with citrate buffer, the tissues were incubated with a primary antibody at 4°C overnight, followed by application of a secondary antibody (Alexa fluor conjugated antibody, 1/1 000 dilutions) for 1 hour at RT. We used the following primary antibodies: ALDH2 (Invitrogen; 1/500 dilution), ANXA1 (OriGene, 1/50 dilution), PRSS3 (Invitrogen, 1/100 dilution), ASS1 (Abcam; 1/100 dilution) and PTX3 (Abcam, 1/100 dilution). All slides were imaged on the Zeiss LSM700 confocal microscope. Three pictures were taken for each tumor section.

### ADDITIONAL RESOURCES

More information can be obtained at <https://clinicaltrials.gov/ct2/show/study/NCT02598219>. ClinicalTrials.gov Identifier: NCT02598219 (Sentirad-1502).