Received: 26 March 2024 | Revised: 5 June 2024 | Accepted: 10 June 2024

DOI: 10.1111/exd.15124

RESEARCH ARTICLE

Experimental Dermatology WILEY

Deciphering dermatological distinctions: Cornulin as a discriminant biomarker between basal cell carcinoma and squamous cell carcinoma detected through e-biopsy and machine learning

Edward Vitkin[1,2] | Julia Wise[1] | Ariel Berl[3] | Ofir Shir-az[3] | Vladimir Kravtsov[4] | Zohar Yakhini[2,5] | Avshalom Shalom[4] | Alexander Golberg[1]

[1]Porter School of Environment and Earth Sciences, Tel Aviv University, Tel Aviv, Israel

[2]Arazi School of Computer Science, Reichman University, Herzliya, Israel

[3]Department of Plastic Surgery, Meir Medical Center, Kfar Sava, Israel

[4]Department of Pathology, Meir Medical Center, Kfar Sava, Israel

[5]Department of Computer Science, Technion—Israel Institute of Technology, Haifa, Israel

Correspondence
Edward Vitkin and Alexander Golberg, Porter School of Environment and Earth Sciences, Tel Aviv University, Tel Aviv, Israel.
Email: edward.vitkin@gmail.com and agolberg@tauex.tau.ac.il

Funding information
TAU's Zimin Center for Technologies for Better Life,; EuroNanoMed 3 MATISSE project; Israel Innovation Authority Kamin Project; TAU's SPARK fund

Abstract

Clinical misdiagnosis between cutaneous squamous cell carcinoma (cSCC) and basal cell carcinoma (BCC) poses treatment challenges and carries risks of recurrence, metastases and increased morbidity and mortality. We aimed to identify discriminant proteins markers for cSCC and BCC using a minimally invasive proteome sampling method called e-biopsy, employing electroporation for non-thermal cell permeabilization and machine learning. E-biopsy facilitated ex vivo proteome extraction from 21 cSCC and 21 BCC pathologically validated human cancers. LC/MS/MS profiling of 126 proteomes was followed by machine learning analysis to identify proteins distinguishing cSCC from BCC. For identified panel validation, we used proteomes sampled by e-biopsy from unrelated 20 cSCC and 46 BCC human cancers, and differential expression analysis of published transcriptomics. Cornulin, the most commonly chosen discriminative biomarker by machine learning models, was also validated using fluorescent immunohistochemistry. One hundred and ninety-two proteomes sampled from one hundred eight patients were analysed. Machine learning-based approaches resulted in a set of 11 potential biomarker proteins that can be used to construct a patient classification model with 95.2% average cross-validation accuracy, BCC precision of 93.6 ± 14.5%, cSCC precision of 98.4 ± 7.2%, specificity of 97.7 ± 11.8% and sensitivity 92.7 ± 15.3%. Protein–protein interaction analysis revealed a novel interaction network connecting 10 of the 11 resulted proteins. Histological and transcriptomic validation confirmed cornulin as a discriminant marker significantly lower in cSCC than in BCC. E-biopsy combined with machine learning provides a novel approach to molecular sampling from skin for biomarker detection and differential expression analysis between cSCC and BCC.

KEYWORDS
cornulin, diagnostics, e-biopsy, electroporation, machine learning, molecular pathology, protein–protein interaction, skin cancer

1 | INTRODUCTION

Cutaneous squamous cell carcinoma (cSCC) and basal cell carcinoma (BCC) are heterogeneous skin lesions that belong to a broad group of nonmelanoma skin cancer (NMSC).[1] Both are increasing in incidence worldwide from 3% to 8% per year since the 1960s.[1] cSCC, representing 20% of all skin cancers, accounts for approximately 75% of all skin cancer deaths, excluding melanoma.[2] BCC, which is the most common malignant neoplasm of humans, is usually curable when the lesion is treated in the early phase.

The clinical diagnoses of cSCC and BCC are performed by a direct skin inspection,[3] often assisted by dermoscopy.[4] In addition, multiple emerging methods such as optical coherence tomography (OCT),[5] reflectance confocal microscopy (RCM),[6] elastic scattering spectroscopy (ESS),[7] and high-frequency ultrasound (HFUS)[8] aim to assist the clinician in diagnosis. Nonetheless, tissue biopsy for histopathological analysis is still essential to confirm the diagnosis, to estimate the risk of recurrence, and to further dictate the treatment pathway.[9]

The clinical features of cSCC and BCC are well described, and, in most cases, the clinical diagnosis is done according to the subsequent histological verification. In some cases, however, these cancers' clinical phenotypes are ambiguous, and discrepancies between the clinical presentations and histologic analyses occur.[10,11] Such misclassification of cSCC as BCC (or vice versa) affects the treatment plan for the lesion.[3] Misclassification of cSCC as BCC carries the highest risk to patients, due to the inherent potential for cSCC's recurrence, metastases and mortality.[12] Recent studies suggested that the introduction of molecular biomarkers, that would accurately stratify non-melanoma skin lesions, would provide a new frontier in personalization of skin cancer care, and hold potential for greatly improving patient diagnosis in many cases. [13] However, direct sampling of the potential biomarkers is still a challenging step, as tissue resection biopsy is limited in the number of sampled sites and has potential side effects.

Recently, we introduced a novel, minimally invasive tissue sampling approach with molecular biopsy using electroporation, termed 'e-biopsy', that selectively extracts liquids with informative proteomes in animal models in liver cancer[14] and brain melanoma[15] in vitro and in breast cancer in vivo, enabling in vivo spatial mapping of differential protein expression.[16] Electroporation-based technologies have been successfully used to permeabilize the cell membrane in vivo, enabling a broad set of applications ranging from tumour ablation to targeted delivery of molecules to cell populations and tissues.[17] However, to the best of our knowledge, such a technology is not yet available and has not been reported in the literature for minimally invasive human tissue molecular sampling. Herein we demonstrate the ability to sample and analyse the proteome extracted ex vivo by e-biopsy from excised human skin tumours. We also show the usefulness of this molecular sampling combined with machine learning methods for potential skin tumour biomarker discovery.

Specifically, we show that proteomic profiles obtained by e-biopsy from cSCC differ from those obtained from BCC. Application of standard machine learning techniques enabled identifying a subset of 11 proteins distinguishing between the two cancers, with the classification performance comparable to current diagnostic methods that involve human professional inspection. One of these proteins, cornulin (CRNN), shows downregulation in cSCC versus BCC as confirmed by immunofluorescent staining and differential expression analysis of the published transcriptomic data. Moreover, proteomic data produced by e-biopsy led to a new protein–protein interaction network jointly covering many of these highly informative proteins. The discovery of such a network together with the cornulin behaviour findings may potentially improve our understanding of systemic molecular differences between the two carcinomas, possibly leading to new target therapies.

2 | METHODS

2.1 | Patients

From March 2020 to May 2024, tissue samples were collected from NMSC (non-melanoma skin cancer) lesions from 108 patients who underwent surgical excision of a skin lesion confirmed pathologically as BCC or cSCC, and additional seven patients samples, confirmed pathologically as benign solar keratosis at Meir Medical Center, Israel. All lesions were excised fresh (10–20 min after surgery), and measured at least 1 cm in diameter. All patients gave written consent for participation and for genetic analysis of tissue. This study was approved by the Meir Medical Center IRB, no. MMC-19-0230.

Malignant samples data were separated into two unrelated cohorts (Figure 1A). Patients from the first cancer cohort, consisting of 42 patients (21 BCC and 21 cSCC, Figure S1), 3 samples per lesion (total 126 samples), were assigned to the initial proteome screening analysis. In addition, out of this batch we selected eight patients (four BCC, four SCC, one sample per lesion) for later immunological staining analysis. Patients from the second cancer cohort, consisting of 66 patients (46 BCC and 20 cSCC), one sample per lesion (total 66 samples), were used for the purpose of validating candidate biomarkers. In total, 192 samples were analysed from 108 tumours from 108 patients (Figure 1A). Solar keratosis patients samples were used for final validation of cornulin expression only.

2.2 | Liquid sample extraction with e-biopsy

The sampling needle is inserted in the sampling location and the ground needle is positioned on the skin surface without penetration, and the pulsed electric fields (PEF) are applied (Figure 1B). A vacuum is applied on the same needle through which the PEF pulses are delivered, to pump the released cellular content into the needle and the syringe. The pulsed electric field was applied using our laboratory custom-made high-voltage pulsed electric field generator, described in detail in Levkov et al.[18] E-biopsy

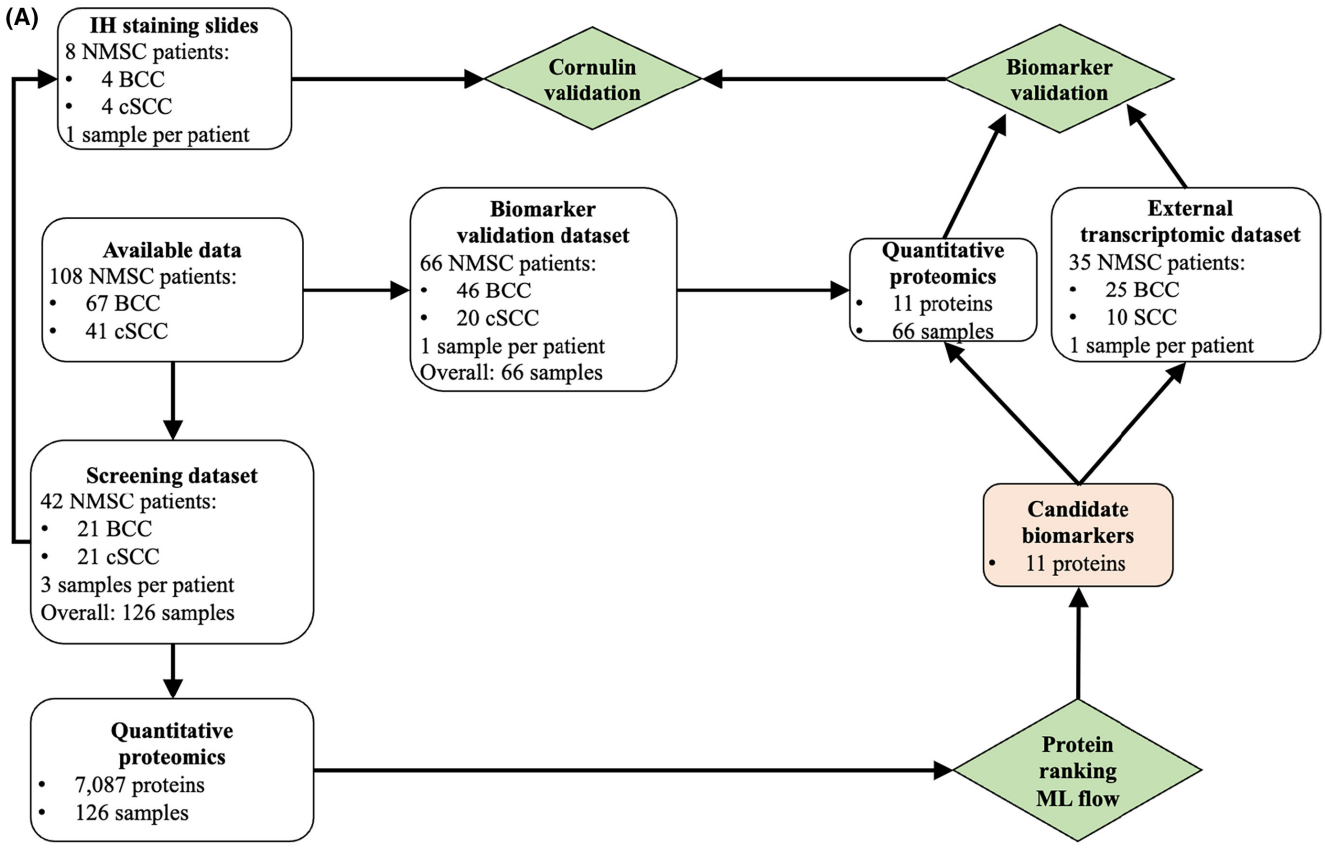

FIGURE 1 (A) Project data flow. (B) Skin e-biopsy procedure ex vivo for discovery and validation of protein signatures differentiating basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC).

was performed using a combination of high-voltage short pulses with low-voltage long pulses[14,19] as follows: 40 pulses at 1000 V, 40 μs, 4 Hz; and 40 pulses at 50 V, 5 ms, delivered at 4 Hz. After the application of the electric fields, the liquids were manually extracted from the tissue to the needle with the vacuum with a 1.5 mL syringe. The liquids were immediately transferred to 1.5 mL tubes with 100 μL double-distilled water. The electrodes were positioned about 5 mm apart. For the sampling electrode, we used a standard 30-G insulin syringe. The second electrode was a custom-made bar measuring 3 mm diameter.

2.3 | Proteins isolation from the e-biopsy sample

Proteins were isolated from the e-biopsy extract using the EZ-RNA II kit (Biological Industries, Ltd., Beit Haemek). Homogenizing solutions were not used in the samples; phase separation solutions were directly added as follows: 0.2 mL of water-saturated phenol, and 0.045 mL of BCP. This step was followed by protein precipitation using isopropanol and wash using guanidine hydrochloride in 95% ethanol. Air-dried protein pellets were taken for proteomic analysis as described below.

2.4 | Identifying and quantifyingf proteins with LC–MS/MS in the proteome screening stage

For the screening stage, we used 63 BCC and 63 cSCC tissue samples from 42 patients (three samples per each patient).

The samples were brought to 8 M urea, 400 mM ammonium bicarbonate, 10 mM DTT, vortexed, sonicated for 5′ at 90% with 10–10 cycles, and centrifuged. The protein amount was estimated using Bradford readings. 20 μg protein from each sample was reduced to 60°C for 30 min, modified with 37.5 mM iodoacetamide in 400 mM ammonium bicarbonate (in the dark, at room temperature, for 30 min), and digested in 2 M Urea, 100 mM ammonium bicarbonate with modified trypsin (Promega) at a 1:50 enzyme: substrate ratio, overnight at 37°C. Additional second digestion with trypsin was done for 4 h at 37°C.

The tryptic peptides were desalted using C18 tips (Harvard Apparatus, MA), dried, and re-suspended in 0.1% formic acid. The peptides were resolved by reverse-phase chromatography on 0.075 × 180-mm fused silica capillaries (J&W Pharmlab, Levittown, PA) packed with Reprosil reversed-phase compound (Dr. Maisch GmbH, Ammerbuch, Germany). The peptides were eluted with a linear 180-min gradient of 5%–28%, a 15-min gradient of 28%–95%, and a 25-min gradient at 95% acetonitrile with 0.1% formic acid in water at flow rates of 0.15 μL/min. Mass spectrometry was performed using a Q-Exactive Plus mass spectrometer (Thermo Fischer Scientific, CA) in a positive mode using a repetitively full MS scan, followed by collision-induced dissociation (HCD) of the 10 most dominant ions selected from the first MS scan. The mass spectrometry data from all of the biological repeats were analysed using MaxQuant software 1.5.2.8 versus the human proteome from the UniProt database with 1% false discovery rate (FDR). The data were quantified by label-free analysis using the same software, based on extracted ion currents (XICs) of peptides, enabling quantitation from each LC/MS/MS run for each peptide identified in any of the experiments.

2.5 | Identifying and quantifying proteins with LC-MS/MS in the biomarker validation stage

For validation of machine learning findings, we used an additional 46 BCC samples and 20 cSCC samples, collected from an additional 66 patients (one sample per patient) and sampled with e-biopsy as described above.

Samples were resuspended in urea buffer (8.5 M urea, 400 mM ammonium bicarbonate, 10 mM DTT). The amount of protein was estimated using Bradford reading. Proteins were digested with trypsin and cleaned on C18 stage tips.

For each of 11 candidate biomarker proteins resulting from the machine learning phase (Table 3) 3–8 peptides were selected according to their abundance among the 126 samples of the screening batch. To extract the current retention time of these peptides, 2 μg from samples that had at least 6 μg of total protein according to the Bradford reading were injected into the mass spec in a regular 3 h gradient MS/MS analysis. All 66 validation batch samples were screened with a 1 min window around the expected peptide retention times.

The resulted mass spectrometry data was analysed using the MaxQuant software version 2.1.1.0 (1) for peak picking and identification using the Andromeda search engine, searching against the human proteome from the UniProt database (from June 2023, 20 605 entries) with mass tolerance of 6 ppm for the precursor masses and 20 ppm for the fragment ions. Oxidation on methionine, lysine and proline and protein N-terminus acetylation were accepted as variable modifications; and carbamidomethyl on cysteine was accepted as static modifications. Minimal peptide length was set to seven amino acids, and a maximum of two miscleavages was allowed. The data were quantified by label-free analysis using the same software. Peptide- and protein-level (FDRs) were filtered to 1% using the target-decoy strategy. The protein results table was filtered to eliminate the identifications from the reverse database.

2.6 | Leveraging machine learning to rank proteins by the contribution to differentiation between cSCC and BCC lesions

Protein signature differentiating cSCC from BCC is effectively a small set of proteins that, when measured together, is informative enough to distinguish between these two conditions. Selection of these most informative proteins was performed by 100 repetitions of leave-6-patient out algorithm (Figure 2), while in each repetition, 18 BCC and 18 SCC patients were used as a training set for protein selection; the remaining three BCC and three SCC patients were used as a testing set to assess the performance of the resulting set of proteins. In each repetition of the leave-6-patient out, the training phase consisted of two major stages: (i) verification of high gene abundance among cSCC or BCC training samples; followed by (ii) iterative backward gene elimination based on relative protein importance. In the first stage, we omitted all proteins with maximal prevalence below 80% among 18 BCC and 18 cSCC training patients; and with maximal prevalence below 50% among 54 BCC and 54 cSCC training samples (each patient was sampled at three locations). With this procedure, we aimed to increase the chances of observing the selected marker in the sample of the corresponding carcinoma type. On average (over differing

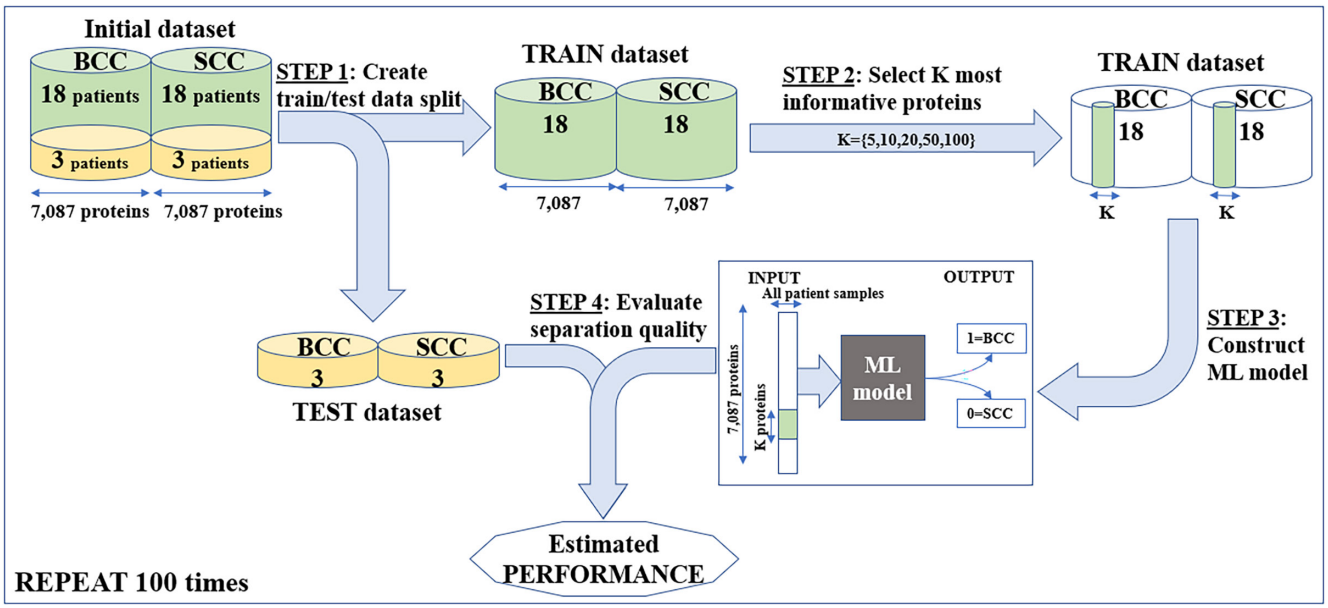

FIGURE 2 General flow of the leave-6-patient out algorithm for the selection of the most informative proteins: (i) splitting the data; (ii) selecting most informative subset of proteins; (iii) construction of machine learning model based on the selected subset; (iv) evaluation of the classification performance of the resulting ML model.

repetitions), roughly 1700 out of 7087 proteins passed this stage in each repetition.

In the second stage, we iteratively omitted 10% of the least important proteins, where the importance of each protein was defined as average protein importance obtained from three different machine learning classifiers: (i) random forest (RF); (ii) logistic regression (LR) and (iii) deep neural network (DNN). Specifically, we used: (i) rank of feature importance for the corresponding protein as obtained from *sklearn* Random Forest classifier; (ii) rank of absolute coefficient for the corresponding protein as obtained from Logistic Regression classifier; and (iii) rank of total absolute weight of the input feature for the corresponding gene in the input layer of Deep Neural Network classifier. The predicted per-patient diagnosis (cSCC patients were classified as positive, and BCC patients as negative) was estimated as the average of predictions of three patients' samples. The resulted model performance was estimated in terms of overall accuracy, positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity over the six patients from the testing set. Finally, all metrics were summarized in terms of mean and its standard error (standard error of mean (SEM) is defined as standard deviation (STD) divided by the square root of the sample size) over 100 repetitions; and 95% Clopper–Pearson confidence interval for the resulted value was estimated (Table 1).

2.7 | Selecting biomarker candidates and evaluating the performance of the summary model

The final set of the most differentiating proteins was derived from the proteins that were consistently (over 100 repetitions of the leave-6-patients-out feature selection algorithm) selected into the final set of the top 5 genes. Specifically, we identified 11 proteins (Table 3) that were chosen with at least 10% frequency (10 reps or more).

The proportion of appearances of each of these proteins was evaluated separately for BCC and for cSCC samples (Figure 3A). The significance of differences between these proportions was estimated by the proportion-based test,[20] which compares the proportions of binarized observations between two groups.

In addition, these proteins were analysed for protein–protein interactions based on IntAct DB,[21] KEGG DB,[22] and other public sources,[23,24] resulting in a protein network describing these and intermediate proteins (Figure 3B).

To evaluate the classification power of this selected subset of proteins, we performed 100 repetitions of leave-6-patient out cross-validation. A Deep Neural Network classifier model was selected as the optimal algorithm for the input of this size based on Table 1. In each repetition, a Deep Neural Network classifier was constructed based on the samples from 18 BCC and 18 cSCC patients, and evaluated on the samples from the remaining three BCC and three cSCC patients. Finally, the model performances were summarized over all repetitions (Table 2).

2.8 | Candidate biomarker validation with additional proteomics and external transcriptomics data

The set of 11 candidate biomarkers (Table 3) was tested with proteomics collected from 66 validation patients (46 BCC and 20 cSCC), one sample per lesion (66 samples total).

TABLE 1 Performance of the best (out of three algorithms) machine learning models selected according to best leave-6-patient out accuracy per each size of protein signature.

No. of used proteins	Machine learning algorithm	Accuracy (correct SCC and BCC predictions out of total cases)	NPV (actual BCC out of predicted BCC)	Specificity (predicted BCC out of total actual BCC)	PPV (actual SCC out of predicted SCC)	Sensitivity (predicted SCC out of total actual SCC)
5	DNN	78.3 ± 1.6% (74.8%–81.6%)	84.2 ± 2.8% (81.9%–91.0%)	66.7 ± 2.9% (61.0%–72.0%)	76.2 ± 1.9% (68.1%–77.4%)	90.0 ± 1.8% (86.0%–93.2%)
10	DNN	81.0 ± 1.7% (77.6%–84.1%)	86.6 ± 2.7% (85.5%–93.6%)	69.7 ± 2.9% (64.1%–74.8%)	78.7 ± 1.8% (70.5%–79.6%)	92.3 ± 1.5% (88.7%–95.1%)
20	DNN	79.5 ± 1.6% (76.0%–82.7%)	88.0 ± 2.5% (84.5%–93.0%)	67.0 ± 2.7% (61.4%–72.3%)	76.5 ± 1.7% (68.8%–78.0%)	92.0 ± 1.6% (88.3%–94.8%)
50	LR	79.2 ± 1.7% (75.7%–82.3%)	90.3 ± 2.6% (89.2%–96.5%)	62.7 ± 2.9% (56.9%–68.2%)	75.0 ± 1.7% (67.2%–76.3%)	95.7 ± 1.2% (92.7%–97.7%)
100	LR	78.5 ± 1.7% (75.0%–81.7%)	88.9 ± 2.8% (88.0%–95.8%)	62.0 ± 2.8% (56.2%–67.5%)	74.1 ± 1.7% (66.7%–75.8%)	95.0 ± 1.4% (91.9%–97.2%)

Note: Per-model details are available in Table S1. All metrics are calculated over 100 repetitions and presented both as mean ± SEM (standard error of mean) and as Clopper–Pearson 95% confidence interval (in parentheses).

Abbreviations: DDN, deep neural network; LR, logistic regression.

For the additional validation, we used an external transcriptomic dataset published by Wan et al.[25] that reports transcript counts per million (CPM) for 16 382 genes in 25 BCC and in 10 SCC patients.

2.9 | Florescent immunohistochemistry

2.9.1 | Slide preparation

The slides (four BCC and four cSCC from eight patients) were immuno-stained for most discriminative biomarker, cornulin (CRNN, UniProt protein ID: Q9UBG3) as follows: Immune florescent staining was performed on 4 μm FFPE sections using the Leica Bond max system (Leica Biosystems, Ltd., Newcastle, UK). Slides were baked at 60°C, dewaxed, and pretreated with epitope-retrieval solutions (ER2, Leica Biosystems) followed by incubation with cornulin antibody (1:100 cat. 11 799 by Proteintech®). Detection was performed using a Donkey anti Rabbit secondary antibody (406 412, Thermo Fisher). To reduce autofluorescence, blocking was performed using the True-View reagent kit (SP-8400 by Vector). Nuclear staining was achieved by incubation with Hoechst 33342 (62 249, Thermo Fisher Scientific). Slides were mounted using VECTASHIELD Vibrance Antifade mounting medium (Vector Labs, Newark, CA).

2.9.2 | Light microscopy photography

Images were taken, using Olympus microscope (BX60, serial no. 7D04032) equipped with a microscope camera (Olympus DP73, serial no. OH05504) at objective magnifications of X10 and X20.

2.9.3 | Histological evaluation

A semi-quantitative analysis of the IF reaction for CRNN was performed using a scoring scale, as follows (X20 field):

Grade 0 = No positive reaction at all.
Grade 1 = Only a few positive cells (<5 cells).
Grade 2 = Very mild immune reaction (5–15 positive cells).
Grade 3 = Mild immune reaction (15–25 positive cells).
Grade 4 = Moderate immune reaction (25–50 positive cells).
Grade 5 = Marked immune reaction (>50 positive cells).

2.9.4 | Immunohistochemistry statistics

Statistical analysis of IF scores was conducted using One-Way Analysis of Variance (ANOVA) followed by Tukey HSD. Significance was considered at $p < 0.05$.

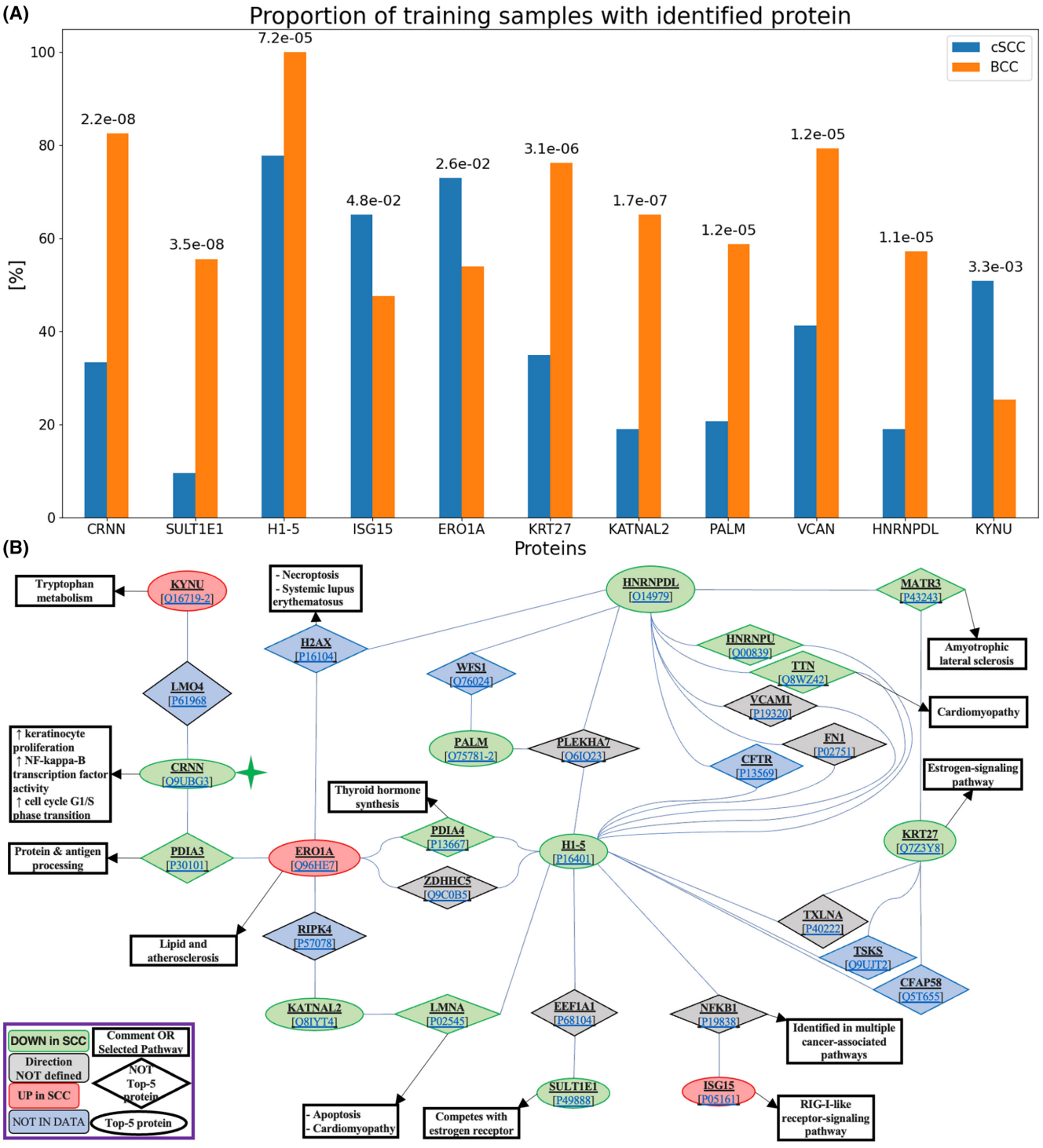

FIGURE 3 (A) Comparative proportion of proteins' (from Table 3) appearance patterns among 126 screening dataset samples. The number above each pair of bars is a proportion-based test[20] *p*-value, quantifying the significance of the difference between two proportions. (B) Protein interaction network of genes (from Table 3) based on IntAct DB.[21] More details on gene appearance, PPI details, and known pathways appear in Tables S2, S3. Shape notation: Ellipses correspond to proteins from Table 3; rhombs correspond to proteins connecting between them; rectangles correspond to pathways and comments. Shape colour: green corresponds to proteins with intensity in cSCC lower than in BCC; red corresponds to proteins with intensity in cSCC higher than in BCC; grey corresponds to proteins without significant change in measured intensity in cSCC versus BCC; blue corresponds to proteins that were not observed in the data. Green star indicates cornulin protein.

TABLE 2 Performance of the initial physician-performed in-clinic diagnoses, and of the summary model, constructed (and summarized over 100 repetitions) on top of 11 biomarker candidates.

Diagnostic setup	Accuracy (% of correct predictions)	BCC precision (real BCC out of predicted BCC)	cSCC precision (real cSCC out of predicted cSCC)	Specificity (predicted BCC out of total real BCC)	Sensitivity (predicted cSCC out of total real cSCC)
Initial in-clinic diagnoses by plastic surgeons pre excision: 14 BCC, 18 cSCC, 10 unclear	61.9% (including unclear cases) 81.2% (excluding unclear cases)	72.2%	92.9%	92.9%	72.2%
Diagnoses with summary model based on 11 identified biomarkers (100 repetitions of leave-6-patients-out cross-validation)	Per sample 88.9 ± 8.1% Per patient 95.2 ± 9.5%	89.9 ± 13.5% 93.6 ± 14.5%	89.7 ± 9.3% 98.4 ± 7.2%	88.2 ± 13.2% 97.7 ± 11.8%	89.6 ± 12.7% 92.7 ± 15.3%

3 | RESULTS

The overall setup of the current research is presented in Figure 1A. The data from 108 non-melanoma skin cancer (NMSC) patients was gathered in two separate batches. First, 126 proteomics samples (a screening dataset) were gathered from 42 NMSC patients, specifically three samples per lesion from 21 BCC and from 21 cSCC patients (Figure S1). Then, a set of 11 potentially interesting biomarker candidates was constructed with the dedicated machine learning flow. At this stage, we gathered an additional (validation) dataset of single-sample-per-lesion proteomic measurements from 46 BCC and 20 cSCC patients, focusing only on these 11 biomarker candidates. Alongside, for the additional validation, we utilized an external transcriptomics dataset published by Wu et al.[25] Finally, the cornulin (CRNN) protein, as the most promising biomarker, was selected and analysed with the immunohistochemistry of four BCC and four cSCC patients.

3.1 | Proteomics harvesting and analysis from excised human skin with e-biopsy

E-biopsy workflow for proteome sampling from freshly (10–20 min since the time of excision) excised human skin is shown in Figure 1B. First, the sampling needle is inserted (0.5–2 mm depth based on skin sample) in the location of interest, the ground needle is positioned on the skin surface (without penetration), and the PEF are applied. Second, a vacuum is applied on the same needle through which the PEF pulses are delivered, to pump the released cellular content into the needle and the syringe. Next, the tissue extract (~1–3 μL) is discharged from the syringe to the external buffer (biology-grade water), and subjected to standard protocols for molecular analysis, including purification, separation, identification (LC/MS/MS in this case), and quantification (MaxQuant). E-biopsy was repeated in three locales in the same area of the excised tissue sample. After this stage, we obtained quantitative measurements of 7087 proteins observed with positive intensity in at least one out of 126 tested samples (on average 1801.6 proteins per sample).

3.2 | Identification of protein signatures that differentiate cSCC and BCC lesions based on e-biopsy-sampled proteome

To identify a group of proteins with maximal potential of distinguishing between cSCC and BCC conditions, we performed 100 repetitions of leave-6-patients out procedure (Figure 2) on the screening dataset of 126 samples from 42 patients (21 BCC and 21 cSCC). In each repetition, 18 cSCC and 18 BCC patients were randomly selected, and their proteomes were used to obtain an optimal subset of 5–100 of the most differentiating proteins. The 18 samples from the remaining three cSCC and three BCC patients were used to assess the classification quality of the resulting protein subset. Finally, for each protein, we counted the number of repetitions (out of 100),

TABLE 3 Proteins selected with frequency above 10% into the final subset of top 5 genes differentiating cSCC and BCC.

Gene symbol	Protein UniProt ID	Gene name	Direction in cSCC versus BCC	No. of repetitions after which the gene was selected into the final top 5
CRNN	Q9UBG3	Cornulin	Down	92/100
SULT1E1	P49888	Sulfotransferase 1E1	Down	65/100
H1-5	P16401	Histone H1.5	Down	41/100
ISG15	P05161	Ubiquitin-like protein ISG15	Up	27/100
ERO1A	Q96HE7	ERO1-like protein alpha	Up	26/100
KRT27	Q7Z3Y8	Keratin, type I cytoskeletal 27	Down	26/100
KATNAL2	Q8IYT4	Katanin p60 ATPase-containing subunit A-like 2	Down	22/100
PALM	O75781	Paralemmin-1	Down	22/100
VCAN*	P13611*	Versican core protein*	Down	19/100
HNRNPDL	O14979	Heterogeneous nuclear ribonucleoprotein D-like	Down	13/100
KYNU	Q16719	Kynureninase	Up	11/100

*VCAN does not appear in Figure 3B.

that led to its being selected for the final set of the top five most differentiating proteins. The most frequently selected proteins, which were selected at least in 10% of repetitions, were further studied for known connections and existing functions.

The resulting average (over 100 repetitions) model accuracy ranges from 69.8% to 81.0% (Table S1), depending upon the specific machine learning algorithm applied (Section 2). Labeling the cSCC subjects positive and BCC subjects negative, we achieved the PPV and sensitivity of 69.8%–78.7% and 79.7%–95.7% respectively, together with NPV and specificity of 73.9%–90.3% and 60.0%–69.7% respectively (Table S1). For each selected size of the protein signature, we present the model with the best performance accuracy (out of three applied machine learning algorithms) in Table 1.

These findings are comparable to the quality of the initial manual diagnostics of the same patients in the clinics (Table 2), as performed by physicians, suggesting that the proposed molecular profiling is potentially a valuable diagnosis-supporting tool, especially if the large number of patients is used for model construction.

Moreover, summarizing the statistics over the proteins selected at the final stage of each repetition (Figure 2), we derived the 11 potentially interesting proteins that appear at least in 10% of repetitions (Table 3). The summary (over 100 additional repetitions) performance of the leave-6-patients out cross-validation based on these 11 proteins only is presented in Table 2. All per-patient metrics clearly outperform current physician-performed in-clinic manual diagnostics, underscoring the diagnostic potential of such a molecular-level classification approach.

3.3 | Proteins from e-biopsy-attained proteome, identified by machine learning models capable of distinguishing BCC from cSCC

Table 3 presents 11 genes that appeared in the top five most useful for constructing the machine learning classifiers in at least in 10% of repetitions of the leave-6-patient out process (Figure 2). Most of these proteins are known to be associated with malignancies, including BCC or cSCC.[26] The analysis of the appearance patterns of these genes shows significant changes between two skin conditions (Figure 3A). Consistently with the ML flow (Section 2), these 11 proteins appear at least in 50% of 63 samples as well as in at least 70% of 21 patients either in BCC or in SCC condition (Figure 3A; Table S4; Figure S2).

Surprisingly, a more detailed study of these 11 proteins revealed that 10 of them comprise a cluster working together. Specifically, the protein–protein interaction data from IntAct DB[21] together with pathways derived from KEGG DB[22] and other public sources[23,24] (Tables S2, S3) provides an interesting protein–protein interaction map for all these proteins, with the exception of VCAN (Figure 3B). The exact network details including (i) protein abundances in our patients, (ii) details and references for experiment that observed each protein–protein interaction, and (iii) known pathways for each protein are presented in Tables S2 and S3.

3.4 | Validation of candidate biomarkers with additional proteomics and external transcriptomics data

For validation purposes, we examined the behaviour of the 11 candidate biomarkers (Table 3) in two additional setups.

First, we quantified the proportion of appearances of these 11 proteins in an additional 66 NMSC patients (46 BCC, 20 SCC, 1 sample per patient, Section 2). We observed (Figure 4A) that all of them, except for KYNU, are aligned (Spearman correlation between proportions is 0.6, *p*-value 3e-3) in the direction of proportion ratio with the proportions observed in the initial screening dataset of 42 patients.

For additional validation, we used an external transcriptomic dataset published by Wan et al.[25] that reports transcript counts

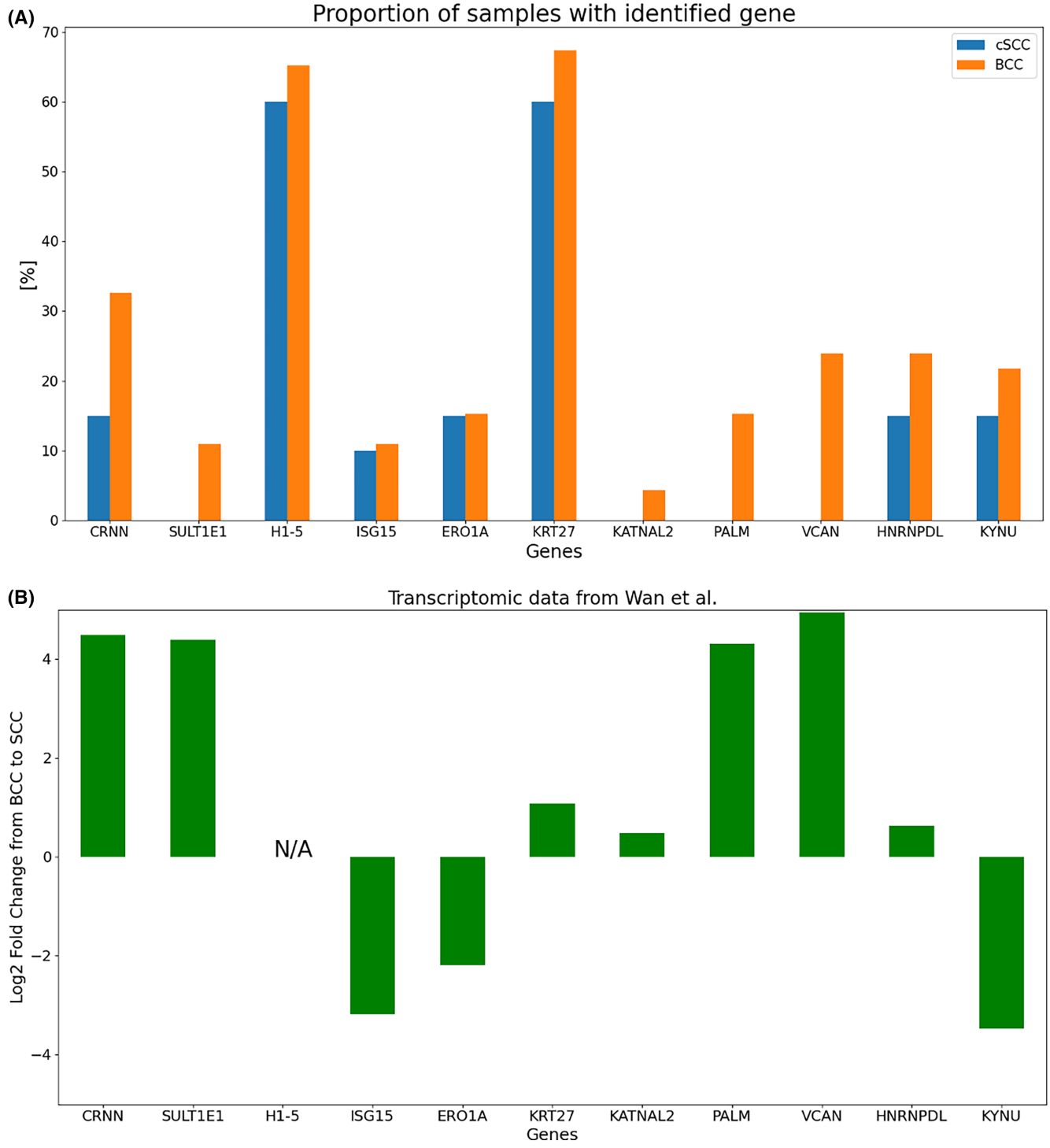

FIGURE 4 Validation of candidate biomarkers from Table 3 with additional data. (A) Proportion of appearances of candidate biomarkers in an additional 66 NMSC patients (Figure 1A). (B) Log2 fold change in average transcript count in Wan et al.[25] transcriptomic dataset.

per million (CPM) for 16 382 genes in 25 BCC and 10 SCC patients. We observed that all 10 successfully mapped proteins (H1-5 was not mapped) align (Spearman correlation between proportions and CPM averages is 0.63, *p*-value 3e-3) in change of direction with the proportions observed in the initial screening dataset of 42 patients.

3.5 | Cornulin as a discriminant biomarker between basal cell carcinoma and squamous cell carcinoma

To additionally test the model predictions, we chose cornulin (CRNN, Uniport ID Q9UBG3), a signalling protein chosen by 92 out of 100 repetitions of the leave-6-patient out process (Figure 2; Table 3). In

Table 4 we summarize all findings on the CRNN over four different validation setups.

In our proteome screening set (Figure 1A), we observed CRNN with positive intensity in 52 out of 63 BCC samples (82.5%), including at least one observation in each of 21 BCC patients (100%); while in cSCC tumour we observed cornulin in only 21 out of 63 samples (33.3%), distributed over 13 out of 21 patients (61.9%), meaning that CRNN is 2.5 times more likely to be observed with positive intensity in BCC than in cSCC (proportion *p*-value of 2.2e-8). Moreover, the average measured intensity among 52 BCC samples with CRNN was 622.2 M, while the average measured intensity among 21 cSCC samples with CRNN was only 135.7 M, meaning that among samples with successfully identified CRNN, its measured intensity is expected to be 4.6 times higher in BCC samples (Mann–Whitney *U* *p*-value of 0.19).

In the biomarker validation proteomic dataset (Figure 1A), the cornulin was observed in 15 (32.6%) of BCC samples with average intensity of 45.3 M versus only 3 (15%) cSCC samples with average intensity of 3.1 M, resulting in a 2:2 ratio of proportions (proportion *p*-value of 0.14) and mean measured positive intensities of 14.5 (Mann–Whitney *U* *p*-value of 0.16).

In the external transcriptomic dataset published by Wan et al.,[25] the average CPM for CRNN protein among 25 BCC patients is 386.6, while among 10 cSCC patients it is 17.2, resulting in a 22.5-fold change ratio (Student *t*-test *p*-value of 8.2e-7) in CRNN transcript counts between BCC and SCC samples.

To additionally validate the e-biopsy sampling approach, we performed an immuno-histochemistry staining for CRNN in four BCC and in four SCC patients selected from the original 42 proteome screening dataset of patients. Intense immuno-staining for CRNN was seen in a distinct part of the BCC tumour (Figure 5). In the SCC tumour, fewer cells showed CRNN expression. The cells' CRNN-positive expression was in the cytoplasm, and appeared granular. The CRNN IF score in the BCC tumours was significantly higher than in the SCC group (4 and 1.25 respectively, where an immunofluorescence score of 4 corresponds to moderate immune reaction (25–50 positive cells) and an immunofluorescence score of 1 corresponds to only a few positive cells (<5 cells)). Finally, we performed differential expression analysis for CRNN for cSCC and BCC subtypes (Figure S3) and in benign samples with solar keratosis (Figure S4, Supplementary information).

4 | DISCUSSION

Besides improving the algorithms for detection of the specific molecular signatures for various diseases, the yet open challenge is biomarker sampling. The current strategy involves molecule extraction using lyses buffers from tissue samples obtained with tissue biopsies. Tissue biopsy procedures carry risks involved with surgical procedures, and may lead to localized tissue injury, bleeding, inflammation, infection and scarring.[27] Moreover, due to these risks, only a few biopsies can be performed in a single procedure, limiting the scope of the spatial mapping of the entire lesion, potentially leading to misdiagnosis if the tumour is only partially sampled, missing critical information due to heterogeneity,[28] or completely missed.[29] To address these problems and to enable direct molecular sampling from tissue without resection, a series of procedures and associated mass spectrometric tools are currently under development,[30] yet none of these to date have obtained broad community recognition. These include the intelligent knife (iKnife), desorption electrospray ionization (DESI),[31] picosecond infrared laser (PIRL), and the MasSpec pen,[30,32] all of which require high energy for sample evaporation, potentially damaging the tissue.[30] Furthermore, the ionization process is competitive, thus, not all informative molecules will be sampled.[30]

In this study, we introduced the potential of e-biopsy, a novel method of molecular sampling with electroporation, for applications in molecular diagnostics of non-melanoma skin cancers. Specifically, we showed that intracellular liquid released by electroporation and directly harvested into a syringe by vacuum, contains a proteome that enables distinguishing cSCC from BCC. We leveraged machine learning feature selection approaches to analyse the sampled proteomes and identify the most promising proteins contributing to such a signature and performed protein–protein interaction analysis of the results.

Cornulin, being the most promising protein in the identified signature, was shown to be differentially expressed in the original and in the additional semi-targeted proteomics dataset, the external transcriptomics dataset, and confirmed by florescent immunohistochemistry. Cornulin, a CRNN protein coding gene, downregulated in cSCC versus BCC, plays a role in epidermal differentiation, while its expression is believed to be specific to squamous cells.[33] It was shown previously that CRNN is downregulated in tongue SCC.[34,35] Moreover, CRNN expression has been reported as differentiating between low-grade and high-grade oral epithelial dysplasia, and could be represented as a potential biomarker for the assessment of progression of oral cancers.[36,37] Up-regulated CRNN levels prevent lesion formation, and its tumour-suppressive role has been reported.[37] Interestingly, a recent histological study of cSCC showed cornulin downregulation in cSCC compared to normal epidermis.[38] Furthermore, cornulin was downregulated in tumours with high histopathological grades when compared to low histopathological-grade tumours.[38] However, its role in basal cells differentiation, to the best of our knowledge is unknown. In addition, its expression over time and various stages/subtypes (Supplementary information, Figures S3, S4) as well as its role in rare lesions such as basosquamous carcinoma requires additional investigation.

In addition to cornulin, other proteins selected by machine learning flow (Table 3) are also potentially interesting biomarker candidates. For example, Sulfotransferase 1E1 (SULT1E1), which also appears in BCC with increased measured intensity compared to cSCC, is responsible for the metabolism of active estrogens and plays crucial roles in their homeostasis.[39] SULT1E1 has been reported as a predictor of breast cancer recurrence[40] and its up-regulation has been reported in normal mammary epithelial cells and breast cancer cell lines.[41–43] SULT1E1 downregulation has

TABLE 4 Details of cornulin behaviour over different validation setups.

Setup name	Setup data	Findings in BCC	Findings in cSCC
Original shotgun proteomics data	Quantitative proteomic measurements of 7087 proteins in: • 21 BCC patients, Three samples per patient • 21 cSCC patients Three samples per patient	• 21 out of 21 BCC patients (100%) • 52 out of 63 BCC samples (82.5%) • Average intensity among these 52 positive samples: 622.2 M	• 13 out of 21 cSCC patients (61.9%) • 21 out of 63 cSCC samples (33.3%) • Average intensity among these 21 positive samples: 135.7 M
Additional shotgun proteomics data	Quantitative proteomic measurements of 11 proteins in: • 46 BCC patients, one sample per patient • 20 cSCC patients, One sample per patient	• 15 out of 46 BCC samples (32.6%) • Average intensity among these 15 positive samples: 45.3 M	• Three out of 20 cSCC samples (15%) • Average intensity among these 3 positive samples: 3.1 M
Transcriptomic dataset published by Wan et al.[25]	Transcript counts per million (CPM) for 16 382 genes in • 25 BCC patients, one sample per patient • 10 SCC patients, one sample per patient	• Average CPM: 386.6	• Average CPM: 17.2
Immuno-histochemistry staining of CRNN	Immuno-florescence score for immune reaction to CRNN antigen in: • 4 BCC patients, one sample per patient • 4 cSCC patients, one sample per patient All 8 patients selected from the 42 original patients	• Immuno-florescence score: 4 * Score 4 corresponds to moderate immune reaction (25–50 positive cells)	• Immuno-florescence score: 1.25 * Score 1 corresponds to only a few positive cells (<5 cells)

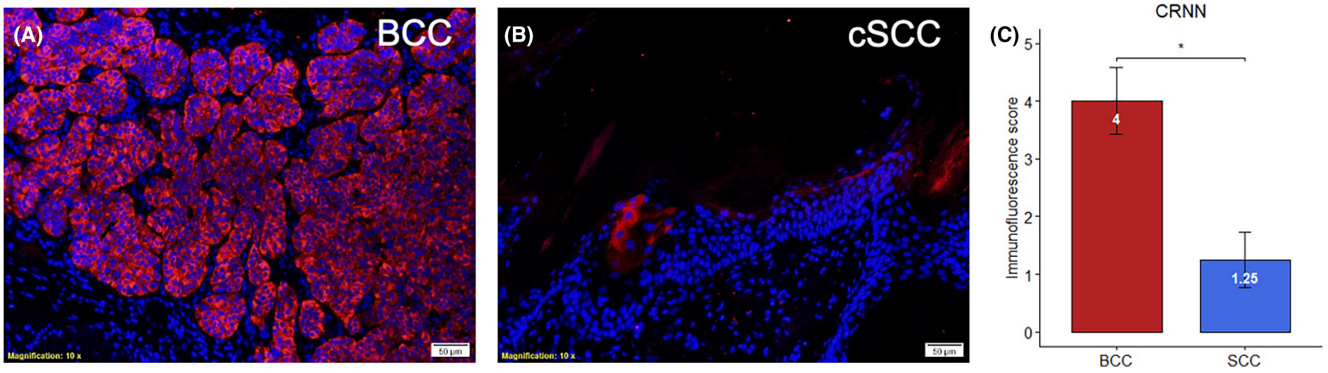

FIGURE 5 Immuno-staining for CRNN (A) Basal cell carcinoma (BCC). (B) Cutaneous squamous cell carcinoma (cSCC). (C) Summary of the semi-quantitative analysis (mean ± SEM) of the IF findings. Statistical analysis was conducted using One-Way Analysis Of Variance (ANOVA) followed by Tukey HSD. Significance is indicated by asterisks: (*) = $p < 0.05$.

been proposed to serve as a marker for more aggressive cancer.[44] Ubiquitin-like protein Interferon Stimulated Gene (ISG15),[45] upregulated in cSCC in our study compared to BCC, was overexpressed in up to 80% of oral squamous cell carcinoma cases, as reported in previous studies.[46] The endoplasmic reticulum oxidoreductin-1-like (ERO1-like protein) alpha, upregulated in cSCC compared to BCC as sampled by e-biopsy proteomes (Figure 3A) is shown in the literature to be associated with cancer,[47] driving the production of VEGF.[48] Interestingly, another study showed significantly higher peritumoral and intra-tumoral blood vessel area in cSCC versus BCC.[49] Versican core protein was previously shown to be a mediator of skin cancer development in mice and humans, and is reported as strongly expressed in BCC compared to cSCC,[26] similar to our findings.

Finally, the protein–protein interaction analysis of these 11 highly informative genes (Table 3) enabled us to identify a novel protein interaction network (Figure 3B), which surprisingly covers 10 of

these 11 top-ranked proteins. Such a network has the potential to be valuable for further understanding of the differences in molecular mechanisms of the two carcinomas.

In this study, we report an expected performance of a binary classifier differentiating between cSCC and BCC based on a limited, predetermined number of proteins (i.e. size of the selected protein signature). The expected per-patient accuracy of a classifier constructed based on the binary observations of 11 highly informative proteins measured over three per-lesion e-biopsy samples is around 95.2% (Table 2). In comparison, the accuracy of the initial diagnosis by dermatologists in clinic on the same 42 patients was only 61.9% (or 81.2% when omitting the unclear cases, Table 2). Indeed, both cSCC precision and specificity in clinic was 92.9% on diagnosed patients, which is higher than current per-sample model performance (89.7% and 88.2%, respectively). One of the major causes of such disparity in metrics can be spatial heterogeneity of the sampled area (i.e. the lesion was not sampled in contrast to the visual observation of the entire lesion region performed by clinician). This can be amended by multiple sampling locations, resulting in 98.4% and 97.7% cSCC per patient precision and specificity respectively for three sampled per-lesion locations (Table 2).

Important to mention that the accuracy of the initial diagnosis in clinic for 42 patients selected for our study does not represent the general distribution of patients diagnostic quality in clinics. For example, an expert panel of pathologists reported 91.6% accuracy for 154 patient panels,[10] and a single surgeon reported accuracy of 93.8% in 1326 cases.[50] Note that both the panel and surgeon reports included significantly higher numbers of BCC patients versus cSCC patients, while in our study, the numbers of cSCC and BCC patients are equal.

One limitation of the current study is its sole focus on the NMSC samples, without addressing its matching healthy areas. Thus, for the next research phase, we are working on gathering healthy tissue samples and their comparative analysis to the available NMSC samples. Moreover, we plan to include paired healthy and affected samples obtained from the same patient to verify that every patient's lesion has a genetically adequate control. Based on proteomic signatures obtained by e-biopsy and the resulting model performance, we aim to develop even better classifiers differentiating cSCC, BCC, and normal skin in future large-scale trials. Such disease-specific classifiers could enable a personalized treatment approach in high-risk NMSC.

AUTHOR CONTRIBUTIONS

Edward Vitkin: Conceptualization, bioinformatics, manuscript drafting and approval. **Julia Wise**: Experiments, protein sampling and analysis. **Ariel Berl**: Experiments, sample collection, pathology, clinics, manuscript review. **Ofir Shir-az**: Experiments, sample collection, pathology, clinics. **Batel Gabay**: Numerical modelling, manuscript drafting. **Amrita Singh**: Experiments, protein sampling and analysis. **Vladimir Kravtsov**: Pathology. **Zohar Yakhini**: Conceptualization, data analysis, critical manuscript review. **Avshalom Shalom**: Conceptualization, critical manuscript review. **Alexander Golberg**: Conceptualization, experiment, data analysis, manuscript drafting.

ACKNOWLEDGEMENTS

The authors thank the Israel Innovation Authority Kamin Project, TAU's SPARK fund, TAU's Zimin Center for Technologies for Better Life, and the EuroNanoMed MATISSE project for their support of this project. All authors thank the members of the Smoler Proteomics Center at the Faculty of Biology at the Technion. We specifically thank Keren Bendalak for her help with the LC–MS/MS analysis.

CONFLICT OF INTEREST STATEMENT

A patent application was filed to protect the electroporation-based sampling technology described herein as invented by AG, JS, EV, AS and ZY.

DATA AVAILABILITY STATEMENT

The authors hereby declare that all data supporting the findings of this study are available in the paper and its Supplementary Information. Data are available via ProteomeXchange with identifier PXD050713.

ORCID

Edward Vitkin https://orcid.org/0000-0003-4055-4772
Ariel Berl https://orcid.org/0000-0003-3705-5454

REFERENCES

1. Crowson AN, Magro CM, Mihm MC. Biopsy interpretation of the skin: primary non-lymphoid cutaneous neoplasia. 2010, 455.
2. Fania L, Didona D, Di Pietro FR, et al. Cutaneous squamous cell carcinoma: from pathophysiology to novel therapeutic approaches. *Biomedicine*. 2021;9:1-33.
3. Mohammad E-A, Mansour M, Parichehr K, Farideh D, Amirhossein R, Ahmad SA. Assessment of clinical diagnostic accuracy compared with pathological diagnosis of basal cell carcinoma. *Indian Dermatol Online J*. 2015;6:258.
4. Sinz C, Tschandl P, Rosendahl C, et al. Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. *J Am Acad Dermatol*. 2017;77:1100-1109.
5. Ferrante di Ruffano L, Dinnes J, Chuchu N, et al. Optical coherence tomography for diagnosing skin cancer in adults. *Cochrane Database Syst Rev*. 2018;12:pp.CD013189-CD013189. https://scholar.google.co.il/scholar?cluster=7806088710657924214
6. Pellacani G, Scope A, Gonzalez S, et al. Reflectance confocal microscopy made easy: the 4 must-know key features for the diagnosis of melanoma and nonmelanoma skin cancers. *J Am Acad Dermatol*. 2019;81:520-526.
7. Rodriguez-Diaz E, Manolakos D, Christman H, et al. Optical spectroscopy as a method for skin cancer risk assessment. *Photochem Photobiol*. 2019;95:1441-1445.
8. Chen ZT, Yan JN, Zhu AQ, et al. High-frequency ultrasound for differentiation between high-risk basal cell carcinoma and cutaneous squamous cell carcinoma. *Skin Res Technol*. 2022;28:410-418.
9. Tanese K. Diagnosis and management of Basal Cell Carcinoma. *Curr Treat Options in Oncol*. 2019;20:1-13.
10. Ryu TH, Kye H, Choi JE, Ahn HH, Kye YC, Seo SH. Features causing confusion between basal cell carcinoma and squamous cell carcinoma in clinical diagnosis. *Ann Dermatol*. 2018;30:64-70.

11. Neagu N, Lallas K, Maskalane J, et al. Minimizing the dermatoscopic morphologic overlap between basal and squamous cell carcinoma: a retrospective analysis of initially misclassified tumours. *J Eur Acad Dermatol Venereol*. 2020;34:1999-2003.
12. Lallas A, Pyne J, Kyrgidis A, et al. The clinical and dermoscopic features of invasive cutaneous squamous cell carcinoma depend on the histopathological grade of differentiation. *Br J Dermatol*. 2015;172:1308-1315.
13. Azimi A, Yang P, Ali M, et al. Data independent acquisition proteomic analysis can discriminate between actinic keratosis, Bowen's disease, and cutaneous squamous cell carcinoma. *J Invest Dermatol*. 2020;140:212-222.e11.
14. Golberg A, Sheviryov J, Solomon O, Anavy L, Yakhini Z. Molecular harvesting with electroporation for tissue profiling. *Sci Rep*. 2019;9:15750.
15. Genish I, Gabay B, Ruban A, et al. Electroporation-based proteome sampling ex vivo enables the detection of brain melanoma protein signatures in a location proximate to visible tumor margins. *PLoS One*. 2022;17:e0265866.
16. Vitkin E, Singh A, Wise J, Ben-Elazar S, Yakhini Z, Golberg A. Molecular Harvesting of Proteins with Electroporation In Vivo Facilitates the Profiling of Spatial Differential Protein Expression in Tumors. 2022. doi:10.21203/RS.3.RS-1242718/V1
17. Yarmush ML, Golberg A, Serša G, Kotnik T, Miklavčič D. Electroporation-based technologies for medicine: principles, applications, and challenges. *Annu Rev Biomed Eng*. 2014;16:295-320.
18. Levkov K, Vitkin E, González CA, Golberg A. A laboratory IGBT-based high-voltage pulsed electric field generator for effective water diffusivity enhancement in chicken meat. *Food Bioprocess Technol*. 2019;12:1993-2003.
19. Ghosh S, Gillis A, Sheviryov J, Levkov K, Golberg A. Towards waste meat biorefinery: extraction of proteins from waste chicken meat with non-thermal pulsed electric fields and mechanical pressing. *J Clean Prod*. 2018;208:220-231. doi:10.1016/J.JCLEPRO.2018.10.037
20. Vitkin E. Differential expression analysis of binary appearance patterns. *Open Res Eur*. 2024;4:52.
21. Orchard S, Ammari M, Aranda B, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. 2014;42:D358-D363.
22. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27-30.
23. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480-D489.
24. Carbon S, Ireland A, Mungall CJ, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics*. 2009;25:288-289.
25. Wan J, Dai H, Zhang X, et al. Distinct transcriptomic landscapes of cutaneous basal cell carcinomas and squamous cell carcinomas. *Genes Dis*. 2021;8:181-192.
26. Kunisada M, Yogianti F, Sakumi K, Ono R, Nakabeppu Y, Nishigori C. Increased expression of versican in the inflammatory response to UVB- and reactive oxygen species-induced skin tumorigenesis. *Am J Pathol*. 2011;179:3056.
27. Abhishek K, Khunger N. Complications of skin biopsy. *J Cutan Aesthet Surg*. 2015;8:239-241.
28. Berl A, Shir-az O, Genish I, et al. Exploring multisite heterogeneity of human basal cell carcinoma proteome and transcriptome. *PLoS One*. 2023;18:e0293744.
29. Wolberink EAW, Pasch MC, Zeiler M, van Erp PEJ, Gerritsen MJP. High discordance between punch biopsy and excision in establishing basal cell carcinoma subtype: analysis of 500 cases. *J Eur Acad Dermatol Venereol*. 2013;27:985-989.
30. Hänel L, Kwiatkowski M, Heikaus L, Schlüter H. Mass spectrometry-based intraoperative tumor diagnostics. *Future Sci OA*. 2019;5:FSO373.
31. Margulis K, Chiou AS, Aasi SZ, Tibshirani RJ, Tang JY, Zare RN. Distinguishing malignant from benign microscopic skin lesions using desorption electrospray ionization mass spectrometry imaging. *Proc Natl Acad Sci USA*. 2018;115:6347-6352.
32. Siu-Yung Chan L. Mohs micrographic surgery sine microscopy: is mass spectrometry an upcoming intraoperative cancer margin assessment tool? *Ann Clin Oncol*. 2019;1–3:1-3. doi:10.31487/J.ACO.2018.01.06
33. Contzler R, Favre B, Huber M, Hohl D. Cornulin, a new member of the 'fused gene' family, is expressed during epidermal differentiation. *J Invest Dermatol*. 2005;124:990-997.
34. Xu Z, Wang MR, Xu X, et al. Novel human esophagus-specific gene c1orf10: cDNA cloning, gene structure, and frequent loss of expression in esophageal cancer. *Genomics*. 2000;69:322-330.
35. Ye H, Yu T, Temam S, et al. Transcriptomic dissection of tongue squamous cell carcinoma. *BMC Genomics*. 2008;9:69. https://scholar.google.co.il/scholar?cluster=16165997153807209632
36. Santosh N, McNamara KK, Beck FM, Kalmar JR. Expression of cornulin in oral premalignant lesions. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2019;127:526-534.
37. Chen K, Li Y, Dai Y, et al. Characterization of tumor suppressive function of cornulin in esophageal squamous cell carcinoma. *PLoS One*. 2013;8(7):1. https://scholar.google.co.il/scholar?cluster=18425705267249689993
38. Karumuri R, Shah D, Arnouk H. Cornulin as a potential novel biomarker for cutaneous squamous cell carcinoma. *Cureus*. 2022;14:e31694. doi:10.7759/cureus.31694
39. Yi M, Negishi M, Lee SJ. Estrogen sulfotransferase (SULT1E1): its molecular regulation, polymorphisms, and clinical perspectives. *J Pers Med*. 2021;11(3):194. https://scholar.google.co.il/scholar?cluster=3714430634190238830.
40. Collin LJ, Cronin-Fenton DP, Ahern TP, et al. Cohort profile: the predictors of breast cancer recurrence (ProBe CaRE) premenopausal breast cancer cohort study in Denmark. *BMJ Open*. 2018;8:21805.
41. Pasqualini JR. Estrogen sulfotransferases in breast and endometrial cancers. *Ann N Y Acad Sci*. 2009;1155:88-98.
42. Shields-Botella J, Chetrite G, Meschi S, Pasqualini JR. Effect of nomegestrol acetate on estrogen biosynthesis and transformation in MCF-7 and T47-D breast cancer cells. *J Steroid Biochem Mol Biol*. 2005;93:1-13.
43. Hasegawa G, Akatsuka K, Nakashima Y, Yokoe Y, Higo N, Shimonaka M. Tamoxifen inhibits the proliferation of non-melanoma skin cancer cells by increasing intracellular calcium concentration. *Int J Oncol*. 2018;53:2157-2166.
44. Xu Y, Lin X, Xu J, Jing H, Qin Y, Li Y. SULT1E1 inhibits cell proliferation and invasion by activating PPARγ in breast cancer. *J Cancer*. 2018;9:1078-1087.
45. Han HG, Moon HW, Jeon YJ. ISG15 in cancer: beyond ubiquitin-like protein. *Cancer Lett*. 2018;438:52-62.
46. Laljee RP, Muddaiah S, Salagundi B, et al. Interferon stimulated gene—ISG15 is a potential diagnostic biomarker in oral squamous cell carcinomas. *Asian Pac J Cancer Prev*. 2013;14:1147-1150.
47. Liu L, Wang C, Li S, et al. ERO1L is a novel and potential biomarker in lung adenocarcinoma and shapes the immune-suppressive tumor microenvironment. *Front Immunol*. 2021;12:2870.
48. Tanaka T et al. Cancer-associated oxidoreductase ERO1-α drives the production of VEGF via oxidative protein folding and regulating the mRNA level. *Br J Cancer*. 2016;114:1227-1234.

49. Pastushenko I, Gracia-Cazaña T, Vicente-Arregui S, et al. Squamous cell carcinomas of the skin explore angiogenesis-independent mechanisms of tumour vascularization. *J Skin Cancer*. 2014;2014:1-5.
50. Shir-Az O, Yaacobi DS, Berl A, Yosefof E, Grush AE, Shalom A. Diagnostic accuracy and cost savings associated with dermoscopy: an economic study. *Semin Plast Surg*. 2022;36:101-106.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1: Supporting information.

How to cite this article: Vitkin E, Wise J, Berl A, et al. Deciphering dermatological distinctions: Cornulin as a discriminant biomarker between basal cell carcinoma and squamous cell carcinoma detected through e-biopsy and machine learning. *Exp Dermatol*. 2024;33:e15124. doi:10.1111/exd.15124