ORIGINAL ARTICLE

Transcriptomic analysis of cutaneous squamous cell carcinoma reveals a multigene prognostic signature associated with metastasis

Jun Wang, PhD,[a] Catherine A. Harwood, MD, PhD,[a,b] Emma Bailey, PhD,[a] Findlay Bewicke-Copley, PhD,[a] Chinedu Anthony Anene, PhD,[a,c] Jason Thomson, MD,[a,b] Mah Jabeen Qamar, BSc, MA,[a,b] Rhiannon Laban, MSc,[a,b] Craig Nourse, PhD,[d] Christina Schoenherr, PhD,[d] Mairi Treanor-Taylor, MD,[d,e] Eugene Healy, MB, PhD,[f,g] Chester Lai, BM, PhD,[f,g] Paul Craig, MD,[h] Colin Moyes, MD,[i] William Rickaby, MD,[j] Joanne Martin, PhD,[a] Charlotte Proby, MD,[k] Gareth J. Inman, PhD,[d,e] and Irene M. Leigh, CBE, DSc[a]

Background: Metastasis of cutaneous squamous cell carcinoma (cSCC) is uncommon. Current staging methods are reported to have sub-optimal performances in metastasis prediction. Accurate identification of patients with tumors at high risk of metastasis would have a significant impact on management.

Objective: To develop a robust and validated gene expression profile signature for predicting primary cSCC metastatic risk using an unbiased whole transcriptome discovery-driven approach.

Methods: Archival formalin-fixed paraffin-embedded primary cSCC with perilesional normal tissue from 237 immunocompetent patients (151 nonmetastasizing and 86 metastasizing) were collected retrospectively from four centers. TempO-seq was used to probe the whole transcriptome and machine learning algorithms were applied to derive predictive signatures, with a 3:1 split for training and testing datasets.

Results: A 20-gene prognostic model was developed and validated, with an accuracy of 86.0%, sensitivity of 85.7%, specificity of 86.1%, and positive predictive value of 78.3% in the testing set, providing more stable, accurate prediction than pathological staging systems. A linear predictor was also developed, significantly correlating with metastatic risk.

Limitations: This was a retrospective 4-center study and larger prospective multicenter studies are now required.

Conclusion: The 20-gene signature prediction is accurate, with the potential to be incorporated into clinical workflows for cSCC. (J Am Acad Dermatol https://doi.org/10.1016/j.jaad.2023.08.012.)

Key words: cutaneous squamous cell carcinoma; machine learning; metastasis; prognosis; risk stratification; transcriptomics.

BACKGROUND

Cutaneous squamous cell carcinoma (cSCC) is the commonest form of skin cancer with metastatic potential and incidence and mortality rising.[1-4] Although the frequency of metastasis arising from cSCC is relatively low at 2% to 5%, the sheer number of cases represents a significant disease burden. Current management could be improved by more accurately identifying tumors most likely to metastasize, targeting adjuvant therapy, and intense clinical supervision programs to those at highest risk, while reducing unnecessary interventions for people with low-risk tumors.

Multiple histopathological staging classifications for cSCC are available although reported to be suboptimal in predicting poor outcomes.[5,6] Recent studies suggest that genomic and transcriptomic signatures may improve risk prediction for primary cSCC progression.[7-10] Using whole exome sequencing data, we previously identified 16 high-risk and 6 low-risk specific significantly mutated genes.[9] More recently, a 40-gene expression profiling (GEP) signature based on candidate genes identified by a combination of literature review and discovery efforts, was developed to predict metastatic risk (Castle Biosciences, Inc Friendswood, Texas).[11,12] A positive predictive value (PPV) of 60% was achieved for the highest risk tumors, with overall sensitivity, specificity, and PPV for differentiating Class 2 (high-) and Class 1 (low-risk) cSCC of 65.4%, 68.8%, and 28.8%, respectively.[11] A completely unbiased discovery-driven approach using information from the whole genome and transcriptome to identify prognostic gene signatures is currently lacking. Such an approach may also uncover key molecular mechanisms underpinning disease progression and metastatic risk.

To develop a validated prognostic signature in an unbiased manner, we assembled a multicentre cohort of primary cSCC archival tissue from 237 patients with known clinical outcomes (no metastasis over 3 years, *n* = 151; metastasis, *n* = 86). Whole transcriptomic data were generated from tumor and perilesional normal skin. A range of machine learning (ML) techniques was applied and a 20-gene GEP model was developed which displayed a

CAPSULE SUMMARY

- A 20-gene expression profile signature derived from clinical archival tissue using an unbiased whole-transcriptome approach showed superior performance for predicting metastatic risks for primary cutaneous squamous cell carcinoma (cSCC).

- This prognostic signature could significantly improve risk stratification, identifying patients with high-risk cSCC who may benefit from adjuvant treatment and reducing overtreatment for patients with low-risk cSCC.

high level of accuracy in differentiating metastasizing and nonmetastasizing primary cSCC. A linear predictor based on the 20-gene GEP was then developed to further aid the implementation of the GEP signature for risk stratification in clinical practice. Ultimately, use of this GEP to guide management decisions may significantly improve patient management for this common cancer.

METHODS

Ethical approval and sample identification

This study was approved as IRAS project 266,559 (Diagnostic marker panel development for progression in skin cancer, PERMEDID). Four collaborating pathology centers identified consecutive patients with primary cSCC which had metastasized, or primary cSCC which had not metastasized within 3 years (Table I). Immunosuppressed patients were excluded. Formalin fixed paraffin embedded (FFPE) sections were reviewed by an expert dermatopathologist and tumor and perilesional normal skin marked for subsequent analysis (see Supplementary Materials, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1).

Pathology review and pathological tumor staging

Hematoxylin and eosin stained sections were digitally scanned by Leica scanner and Aperio software. Images were reviewed centrally by two expert dermatopathologists and primary tumors typed, graded, and histologically staged using Union for International Cancer Control (UICC)-8 and Brigham and Women's Hospital (BWH) classifications.

Transcriptomics investigation

Transcriptomic analysis was performed using the TempO-Seq whole-protein coding transcriptome platform with a proprietary processing pipeline (Bioclavis Ltd, Glasgow, UK).[13] Data preprocessing and normalization were performed using limma R package.[14] Batch effect was removed using the ComBat package.[15] Differential expression (DE) analysis using limma was performed between clinical groups, followed by gene set overrepresentation

---

Abbreviations used:

AUC:	area under the ROC curve
BCR:	B-cell receptor
BWH:	brigham and women's hospital
cSCC:	cutaneous squamous cell carcinoma
DE:	differential expression
ESCC:	esophageal squamous cell carcinoma
FFPE:	Formalin fixed paraffin embedded
GEP:	gene expression profile
GSEA:	gene set enrichment analysis
KNN:	K-nearest neighborhood
ML:	machine learning
PPV:	positive predictive value
UICC:	union for international cancer control

---

and gene set enrichment analysis (GSEA) using DAVID[16] and clusterProfiler.[17]

Gene signature analysis using machine learning

To derive a set of genes that could distinguish two groups (ie, metastasizing vs nonmetastasizing cSCC), the caret R package[18] was used for ML analysis. A range of ML techniques were used and compared (Supplementary Materials). We randomly split the samples into training (75%) and testing (25%) sets. Starting with an initial set of genes in the training set (ie all DE genes from the DE analysis comparing metastasizing and nonmetastatic cSCC), the best performing set of genes for each ML algorithm (ie, feature selection) was determined using the Recursive Feature Elimination procedure, with 10-fold repeated cross validation of five repeats. A final model for each ML algorithm was then trained using the final selected number of genes with 10-fold repeated cross validation of ten repeats and used to predict the two classes in the testing set. The performances of predictions were measured using accuracy, precision, along with sensitivity and specificity, PPV, and negative predictive value.

A weighted linear predictor was generated for each sample based on the expression of the final set of genes in the model and their fold changes in the DE analysis (see Supplementary Materials) Linear predictors were compared between clinical groups and correlated with classes. The area under the ROC curve (AUC) was calculated using the pROC package.[19]

RESULTS
Clinicopathologic characteristics

Demographic details of patients and histologic features of primary cSCC are presented in Table I.

Transcriptomic analysis between primary cSCC groups

GEP of 19,072 genes across a total of 433 samples were sufficiently profiled for analysis. Four sample groups were compared; cSCC tumor from metastasizing ($n = 84$) and nonmetastasizing ($n = 146$) cSCC, and matched perilesional normal skin from metastasizing ($n = 71$) and nonmetastasizing ($n = 132$) cSCC (Supplementary Table I, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1). Principal component analysis based on genes across all samples showed a clear separation between cSCC and perilesional normal skin samples from metastasizing and non-metastasizing cSCC (Supplementary Fig 1, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1). Differential gene expression analysis revealed that 1038 genes were upregulated and 236 genes downregulated in metastasizing cSCC compared to nonmetastasizing cSCC (absolute $\log_2$ fold change >1 and adjusted P value $<.05$). The gene set overrepresentation test showed keratinization, B-cell receptor (BCR), innate immune response, cell cycle, DNA replication, and DNA repair were highly overrepresented in the DE genes (hypergeometric test $q < 0.05$, Supplementary Fig 2A, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1). Overrepresentation analysis against cellular signatures showed that signatures associated with neural progenitor, endothelial, and cancer stem cells were highly enriched within the DE genes (Supplementary Fig 2B, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1), suggesting that cell differentiation is a key factor distinguishing the two cSCC groups. GSEA against MSigDB canonical pathways further suggested that cell cycle related, DNA replication and repair, and immune pathways (BCR regulation, interferon, and interleukin-12 signaling), were all significantly upregulated in metastasizing cSCC, while formation of the cornified envelope, keratinization, and many metabolism pathways (sphingolipid, triglyceride, creatine, and fatty acid metabolism) were significantly downregulated (Fig 1).

Normal perilesional samples from metastasizing and nonmetastasizing primary cSCC were also compared. GSEA indicated many immune pathways (such as BCR and T cell receptor signaling, Fc gamma receptor activation, and chemokine receptor binding) and cell cycle related pathways (synthesis, replication and repair of DNA) were significantly upregulated in perilesional skin samples from metastasizing tumors (Supplementary Table II, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1).

Table I. Clinicopathologic details of patients and primary cSCC samples

Feature	All (n = 237)	No metastasis (n = 151)*	Metastasis (n = 86)[†]	P value
Age, y, median (range)	80 (39-100)	78 (39-100)	80 (64-93)	.57
Male, n (%)	142 (60)	90 (60)	52 (60)	1
Located on head and neck, n (%)[‡]	155 (65)	91 (61)	64 (74)	.033
Tumor diameter, cm, mean (range)[§]	1.85 (0.18-9)	1.31 (0.18-4.1)	2.82 (1.6-9)	<.0001
Tumor thickness, mm, mean (range)[‖]	3.94 (0.2-26.7)	2.96 (0.2-13)	5.65 (0.3-26.7)	<.0001
Poorly differentiated, n (%)	115 (48.3)	47 (30.9)	68 (79.1)	<.0001
Clark level > V (beyond fat), n (%)[¶]	43 (18.6)	10 (6.7)	33 (40.2)	<.0001
PNI, n (%)[#]				.0004
Present (≥ 0.1 mm)	20 (8.6)	8 (5.3)	12 (14.6)	
Present (<0.1 mm or unknown)	11 (4.7)	3 (1.99)	8 (9.8)	
Not present	202 (86.7)	140 (92.7)	62 (75.6)	
Lymphovascular invasion**	15 (6.5)	1 (0.66)	14 (17.5)	<.0001
UICC T stage, n (%)[††]				<.0001
T1	134 (59.3)	115 (78.8)	19 (23.75)	
T2	25 (11.1)	11 (7.5)	14 (17.5)	
T3	67 (29.6)	20 (13.7)	47 (58.75)	
T4	-	-	-	
BWH T stage, n (%)[§§]				<.0001
T1	86 (37.7)	84 (56.75)	2 (2.5)	
T2a	65 (28.5)	44 (29.7)	21 (26.25)	
T2b	71 (31.1)	20 (13.5)	51 (63.75)	
T3	6 (2.6)	-	6 (7.5)	

*Total number of primary cSCC which did not metastasize = 152 (one patient had 2 separate primary cSCCs); median follow-up was 76 months.

[†]Median time from primary cSCC to metastasis was 9.9 months.

[‡]Location not recorded for 2 cSCCs (both nonmetastasizing).

[§]Not available for 10 cSCC (5 nonmetastasizing and 5 metastasizing).

[‖]Not available for 15 cSCC (10 nonmetastasizing and 5 metastasizing).

[¶]Invasion through or beyond subcutaneous fat: not available for 7 cSCC (3 nonmetastasizing and 4 metastasizing).

[#]Not available for 5 cSCC (1 nonmetastasizing and 4 metastasizing cSCC).

**Lymphovascular invasion not available for 6 cSCC (all metastasizing).

[††]Not available for 12 cSCC (6 non metastasizing and 6 metastasizing).

[§§]Not available for 10 cSCC (4 nonmetastasizing and 6 metastasizing).

Development of the 20-GEP prognostic signature

To identify a smaller set of genes that were predictive for primary cSCC metastasis, a range of ML classification algorithms were applied after splitting the primary cSCC samples into training and validation sets. A 20-gene model derived from K-nearest neighbors (KNN) was identified (Supplementary Table III, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1) which provided the best performance in differentiating the two cSCC groups in the validation set (n = 57: 36 nonmetastasizing; 21 metastasizing), with an accuracy of 86.0 (95% confidence interval 74.2%-93.7%), a sensitivity of 85.7% and a specificity of 86.1% (Table II). Patients predicted as high risk of metastasis by the 20-GEP signature (n = 23) had significantly worse metastasis-free survival (MFS) rates than those predicted as low-risk (n = 34) (3-year MFS, 91.7% for low-risk vs 21.7% for high-risk) (Fig 2). In this 20-gene GEP model, 18 genes

were upregulated in nonmetastasizing cSCC and 2 genes (*MDK* and *STMN1*) were upregulated in metastasizing cSCC (Supplementary Table III and Fig 3, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1). Functional annotation of the 20 genes suggested the significant enrichment in the signatures from keratinization, GnRH, oxytocin, Ras, and MAPK signaling pathways (hypergeometric test, *P* < .01). Using the same ML procedure based on perilesional normal skin samples, a 22-gene KNN model was also developed with an accuracy of 64.0% (95% CI: 49.2% to 77.1%), sensitivity of 41.2%, and specificity of 75.8% (Table II).

Prognostic accuracy of the 20-GEP test compared to pathological staging classifications

Using the Royal College of Pathologists dataset for histopathological reporting of primary invasive cSCC, tumors were staged by both UICC-8 TNM, and BWH T-staging classifications after central
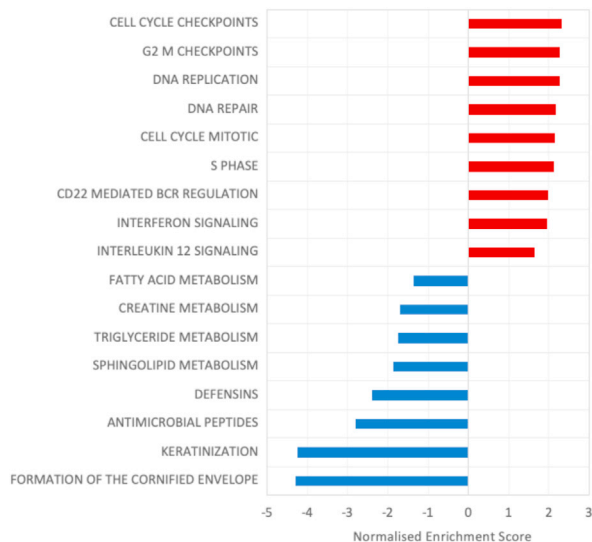
Fig 1. Normalized enrichment scores (*NES*) of the top dysregulated canonical pathways between metastasizing and nonmetastasizing cSCC. Pathways with positive NES (in *red*) were upregulated while pathways with negative NES (in *blue*) were downregulated in metastasizing compared to non-metastasizing primary cSCC.

consensus histopathological review. Prognostic metrics for UICC-8 (low T1/T2 vs high T3/T4) and BWH (low T1/T2a vs high T2b/T3) staging showed performance with an accuracy of 85.4% for both systems in the validation set, compared to 86.0% for the 20-GEP signature (Table II). Performance of BWH T-staging based on original pathology reports without central consensus review (BWH v1), was marginally inferior in predicting metastasis, with an accuracy of 81.8%. This was largely due to differences between the scoring of poor differentiation after central review compared to the original report (Table I, Supplementary Table IV, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1).

The 20-GEP signature showed strong correlations with staging for risk prediction in the validation set. Of 23 metastasizing cases predicted by the 20-GEP test, 21/23 (91.3%) were T2b/T3 by BWH staging vs 15/23 (65.2%) UICC-8 T3/4. Of 32 nonmetastasizing cSCC predicted by the 20-GEP, 26/32 were T1/T2a by BWH and 26/32 were UICC-8 T1/T2 (81.3%). Accuracy of the histology staging systems dropped to 81.1% and 76.5% for BWH and UICC8, respectively, when the whole cohort (*n* = 237) was considered (Table II).

Generation of a linear predictor for metastatic prediction

To further enhance the potential clinical application of the 20-GEP signature, a linear predictor for

metastasis combining the expression values and fold-changes of these 20 genes in the DE analysis was generated: the higher the linear predictor value, the higher the risk of developing metastasis. The previously reported 40-GEP[11] stratifies tumors into 3 classes of risk (low, high, highest), whereas a linear predictor allows a more detailed assessment of risk that can be used alongside pathological risk factors to influence clinical management. The linear predictor had a very high correlation with metastatic risk, with an AUC of 0.85 (95% CI, 0.80-0.91) and 0.88 (95% CI, 0.78-0.99) for the training and validation (testing) sets, respectively (Fig 3). In comparison, the KNN binary classification model (ie, yes or no for metastasis prediction) had an AUC of 0.86 (0.76-0.96). As expected, the linear predictor was significantly higher in metastasizing vs nonmetastasizing cSCC in both training and testing sets (Wilcoxon rank sum test, $P < .0001$, Supplementary Fig 4, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1).

Finally, the linear predictors across both tumor and perilesional skin for both metastasizing and nonmetastasizing cSCC were compared (Supplementary Fig 5, available via Mendeley at https://doi.org/10.17632/z77kdgddm9.1). There was no difference in linear predictors between nonmetastasizing cSCC and both normal adjacent groups. However, linear predictors increased significantly for metastasizing cSCC compared to other groups ($P < .0001$), suggesting that our linear predictor was only associated with metastasizing primary tumors.

DISCUSSION

This study reports a 20-GEP signature that predicts metastatic risk of primary cSCC. It was developed and validated in a UK cohort of 237 primary cSCC from immunocompetent individuals using archival FFPE tissue in which whole-transcriptome analysis with an unbiased discovery approach was performed. The 20-GEP signature achieved an accuracy of 86.0%, a negative predictive value of 91.2% and a PPV of 78.3% for predicting metastasis in the validation set (*n* = 57). A linear predictor to facilitate potential clinical use of the 20-GEP was created based on the expression and fold changes of signature genes and had an AUC of 0.88. UICC-8 TNM and BWH pathological staging systems performed unexpectedly well in risk prediction compared with previous reports. Nonetheless, the 20-GEP remained overall the most stable and accurate predictor of metastatic risk, and in contrast to histology, the GEP signature is unbiased and not dependent on human evaluation and interpretation.

Table II. Accuracy of the prediction of metastatic risks of the 20-GEP signature and other risk assessment methods (*n* = 57)

Classifier	Accuracy%	Sensitivity%	Specificity%	PPV%	NPV%	+LR	-LR
20-GEP	86.0	85.7	86.1	78.3	91.2	6.17	0.17
UICC-8	85.4	81.0	88.2	81.0	88.2	6.88	0.22
BWH	85.4	95.2	79.4	74.1	96.4	4.63	0.06
BWH v1	81.8	76.2	85.3	76.2	85.3	5.18	0.28
22-GEP*	64.0	41.2	75.8	46.7	71.4	1.70	0.78
UICC-8[†]	76.5	58.8	86.3	70.1	79.2	4.29	0.48
BWH[†]	81.1	71.2	86.5	74.0	84.8	5.27	0.33

+LR, Positive likelihood ratio; *-LR*, negative likelihood ratio; *BWH*, brigham and women's hospital staging system after the central review; *BWH v1*, derived from original pathology reports before central pathology review; *GEP*, gene expression profile; *NPV*, negative predictive value; *PPV*, positive predictive value; *UICC*, union for international cancer control.
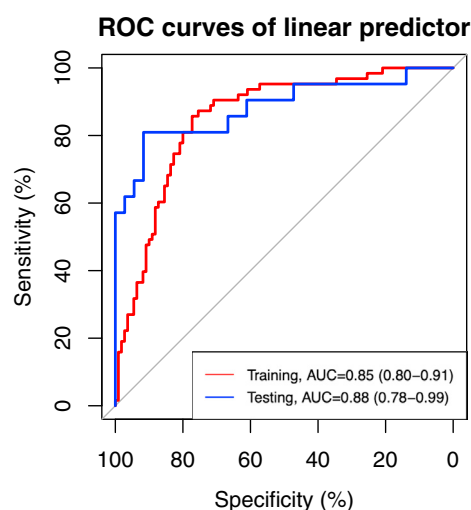*22-GEP was derived from normal adjacent samples only.
[†]Statistics were derived from the whole cohort (*n* = 237).
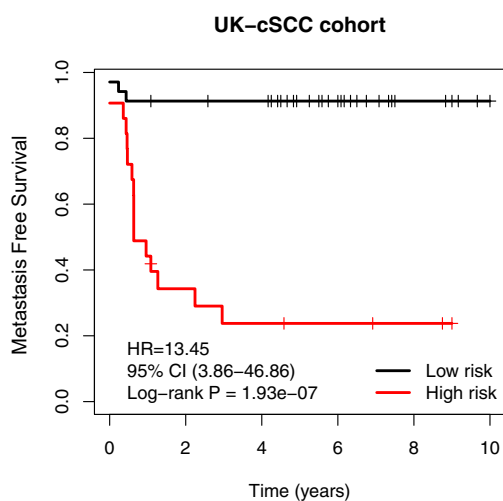
Fig 2. Kaplan-Meier analysis of the 20-GEP prognostic test and outcomes in terms of metastasis free survival in the validation dataset. No at-risk in the follow-up was shown in the table below.

Fig 3. Area under the receiver operating characteristic curve (*AUC*) of the performance of linear predictors correlating with the metastatic incidences. Linear predictors were produced based on the 20-GEP signature, and both training and testing data sets were included in the calculation. AUC and 95% confidence interval were shown.

There appeared to be a strong association between the 20-GEP and keratinization. Key keratinization genes, such as *LCE1C, LCE2B/C, LCE3C,* and *CDSN,* were all significantly downregulated in metastasizing primary cSCC as were two genes involved in alpha-Linolenic acid and ether lipid metabolism (*PLA2G4E/F*), consistent with our GSEA results. Only two genes, *STMN1* and Midkine (*MDK),* were significantly upregulated in metastasizing samples. STMN1 a microtubule-destabilizing protein, regulates the dynamics of microtubules and cell cycle progress.[20] Its high expression is associated with poor prognosis in esophageal (ESCC), lung (lung squamous cell carcinoma) and oral SCC.[21-23] In ESCC

and lung squamous cell carcinoma, it was reported to promote cell proliferation, migration, chemoradiation resistance,[21,22,24] and is strongly associated with lymph node metastasis in ESCC.[25,26] MDK, a heparin-binding growth factor, is also associated with cancer progression, drug resistance, and a tolerogenic and immune-resistant state.[27-30] A recent study showed that MDK was highly expressed by stem-like tumor cells and led to mTOR inhibition persistence and an immune-suppressive microenvironment.[31] MDK represents an interesting therapeutic target for advanced cSCC.

Currently, clinical pathways determining treatment plans for patients with cSCC use

clinicopathological staging systems. In practice, the predictive accuracy of staging systems for primary cSCC can vary significantly across reported studies.[11,32-35] Factors possibly accounting for the variability in pathology staging include nonstandardized reporting of high-risk features (particularly poor differentiation and perineural invasion); problems defining the state of differentiation of an individual tumor; and variable practice in the use of Mohs' surgery which may affect detection of high-risk features and lead to understaging.[11] In our study, careful central review by two highly experienced dermatopathologists adhering to the Royal College of Pathologists dataset led to a much higher performance of pathology staging systems than previously published. This highlights the need for a more objective grading system such as that used worldwide in breast carcinoma.[36]

Additional strengths of our study include an unbiased discovery-driven approach using the whole transcriptome of FFPE clinical samples to develop a prognostic signature suitable for routine clinical use. We also excluded immunosuppressed patients as iatrogenic and disease-associated immunosuppression is an important risk factor for poor outcomes in cSCC and variations in immune status and effects of immunosuppressive drugs are likely to impact the transcriptome. Excluding confounding factors due to immunosuppression may have permitted generation of a more metastasis-specific gene signature of greater use for risk prediction. More work is needed to test our 20-gene signature in other patient populations, such as those with darker skin and in immunosuppressive populations.

The retrospective nature of this study was a limitation and, although consecutive eligible primary cSCC were enrolled at each center, the possibility of some bias relating to patient and sample selection cannot be excluded. The study size for the validation set was also a limitation and further validation will require larger, prospective studies.[5]

In conclusion, we have used an unbiased discovery-driven approach to generate a promising candidate 20-GEP prognostic signature for cSCC metastasis. The GEP not only represents a novel and potentially clinically applicable prognostic tool but has also provided biological insights into the process of metastasis and potential therapeutic targets. In addition, there are biological and genomic mechanisms common to cSCC across different tissue types and this signature may provide further insights into common differentiation and stem-like pathways underpinning these SCCs. Further prospective evaluation is now underway to confirm clinical utility of this GEP in management of primary cSCC.

Conflict of interest

Dr Harwood is honoraria for advisory boards from Sanofi/Regeneron, Almirall, AmLo Biosciences, Incanthera, Leo Pharma, L'Oreal. Other authors have no disclosures.

REFERENCES

1. Venables ZC, Nijsten T, Wong KF, et al. Epidemiology of basal and cutaneous squamous cell carcinoma in the U.K. 2013-15: a cohort study. *Br J Dermatol*. 2019;181(3):474-482.
2. Venables ZC, Autier P, Nijsten T, et al. Nationwide incidence of metastatic cutaneous squamous cell carcinoma in England. *JAMA Dermatol*. 2019;155(3):298-306.
3. Kwiatkowska M, Ahmed S, Ardern-Jones MR, et al. A summary of the updated report on the incidence and epidemiological trends of keratinocyte cancers in the UK 2013-2018. *Br J Dermatol*. 2022;186(2):367-369.
4. Detailed statistics from the get data out (>skin tumours) programme, National Disease Registration Service. Accessed August 31, 2023. https://www.cancerdata.nhs.uk/getdataout/skin
5. Venables ZC, Tokez S, Hollestein LM, et al. Validation of four cutaneous squamous cell carcinoma staging systems using nationwide data. *Br J Dermatol*. 2022;186(5):835-842.
6. Tokez S, Venables ZC, Hollestein LM, et al. Risk factors for metastatic cutaneous squamous cell carcinoma: refinement and replication based on 2 nationwide nested case-control studies. *J Am Acad Dermatol*. 2022;87(1):64-71.
7. South AP, Purdie KJ, Watt SA, et al. Prevalent NOTCH1 mutations in squamous cell carcinogenesis are an early event. *J Invest Dermatol*. 2014;134:2630-2638.
8. Cammameri P, Rose AM, Vincent DF, et al. Inactivation of TGFb receptors in stem cell drives cutaneous squamous cell carcinoma. *Nature Commun*. 2016;25:12493.
9. Inman GJ, Wang AJ, Ai N, et al. The genomic landscape of cutaneous squamous cell carcinoma from immunosuppressed and immunocompetent patients reveals common drivers and a novel mutational signature associated with chronic azathioprine exposure. *Nat Commun*. 2018;9(1):3667.
10. Thomson J, Bewicke-Copley F, Anene CA, et al. The genomic landscape of actinic keratosis. *J Invest Dermatol*. 2021;141(7):1664-1674.e7.
11. Wysong A, Newman JG, Covington KR, et al. Validation of a 40-gene expression profile test to predict metastatic risk in localized high-risk cutaneous squamous cell carcinoma. *J Am Acad Dermatol*. 2021;84(2):361-369.

12. Ibrahim SF, Kasprzak JM, Hall MA, et al. Enhanced metastatic risk assessment in cutaneous squamous cell carcinoma with the 40-gene expression profile test. *Future Oncol*. 2022;18(7): 833-847. https://doi.org/10.2217/fon-2021-1277

13. Yeakley JM, Shepard PJ, Goyena DE, VanSteenhouse HC, McComb JD, Seligmann BE. A trichostatin a expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLoS One*. 2017;12(5):e0178302.

14. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

15. Leek JT, Johnson WE, Parker HS, Sva, et al. Surrogate Variable Analysis. 2022 R Package Version 3.44.0. *Bioconductor*. 2022.

16. Huang Da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.

17. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3):100141.

18. Kuhn M. Building predictive models in R using the caret package. *J Stat Software*. 2008;28:1-26.

19. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12:77.

20. Rubin CI, Atweh GF. The role of stathmin in the regulation of the cell cycle. *J Cell Biochem*. 2004;93(2):242-250.

21. Suzuki S, Yokobori T, Altan B, et al. High stathmin 1 expression is associated with poor prognosis and chemoradiation resistance in esophageal squamous cell carcinoma. *Int J Oncol*. 2017;50(4):1184-1190.

22. Bao P, Yokobori T, Altan B, et al. High STMN1 expression is associated with cancer progression and chemo-resistance in lung squamous cell carcinoma. *Ann Surg Oncol*. 2017;24(13): 4017-4024.

23. Ma HL, Jin SF, Tao WJ, Zhang ML, Zhang ZY. Overexpression of stathmin/oncoprotein 18 correlates with poorer prognosis and interacts with p53 in oral squamous cell carcinoma. *J Craniomaxillofac Surg*. 2016;44(10):1725-1732.

24. Ni PZ, He JZ, Wu ZY, et al. Overexpression of Stathmin 1 correlates with poor prognosis and promotes cell migration and proliferation in oesophageal squamous cell carcinoma. *Oncol Rep*. 2017;38(6):3608-3618.

25. Li J, Qi Z, Hu YP, Wang YX. Possible biomarkers for predicting lymph node metastasis of esophageal squamous cell carcinoma: a review. *J Int Med Res*. 2019;47(2):544-556.

26. Jiang W, Huang S, Song L, Wang Z. STMN1, a prognostic predictor of esophageal squamous cell carcinoma, is a marker of the activation of the PI3K pathway. *Oncol Rep*. 2018;39(2): 834-842.

27. Filippou PS, Karagiannis GS, Constantinidou A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. *Oncogene*. 2020;39(10):2040-2054.

28. Yu X, Zhou Z, Tang S, et al. MDK induces temozolomide resistance in glioblastoma by promoting cancer stem-like properties. *Am J Cancer Res*. 2022;12(10):4825-4839.

29. Olmeda D, Cerezo-Wallis D, Riveiro-Falkenbach E, et al. Whole-body imaging of lymphovascular niches identifies pre-metastatic roles of midkine. *Nature*. 2017;546(7660):676-680.

30. Cerezo-Wallis D, Contreras-Alcalde M, Troulé K, et al. Midkine rewires the melanoma microenvironment toward a tolerogenic and immune-resistant state. *Nat Med*. 2020;26(12):1865-1877.

31. Tang Y, Kwiatkowski DJ, Henske EP. Midkine expression by stem-like tumor cells drives persistence to mTOR inhibition and an immune-suppressive microenvironment. *Nat Commun*. 2022;13(1):5018.

32. Jambusaria-Pahlajani A, Kanetsky PA, Karia PS, et al. Evaluation of AJCC tumor staging for cutaneous squamous cell carcinoma and a proposed alternative tumor staging system. *JAMA Dermatol*. 2013;149(4):402-410. https://doi.org/10.1001/jamadermatol.2013.2456

33. Karia PS, Jambusaria-Pahlajani A, Harrington DP, Murphy GF, Qureshi AA, Schmults CD. Evaluation of American joint committee on cancer, international union against cancer, and Brigham and women's hospital tumor staging for cutaneous squamous cell carcinoma. *J Clin Oncol*. 2014;32(4):327-334. https://doi.org/10.1200/JCO.2012.48.5326

34. Karia PS, Morgan FC, Califano JA, Schmults CD. Comparison of tumor classifications for cutaneous squamous cell carcinoma of the head and neck in the 7th vs 8th edition of the AJCC Cancer Staging Manual. *JAMA Dermatol*. 2018;154:175-181.

35. Ruiz ES, Karia PS, Besaw R, Schmults CD. Performance of the American joint committee on cancer staging manual, 8th edition vs the Brigham and women's hospital tumor classification system for cutaneous squamous cell carcinoma. *JAMA Dermatol*. 2019;155(7):819-825. https://doi.org/10.1001/jamadermatol.2019.0032

36. Blamey RW, Ellis IO, Pinder SE, et al. Survival of invasive breast cancer according to the Nottingham prognostic index in cases diagnosed in 1990–1999. *Eur J Cancer*. 2007;43:1548-1555.