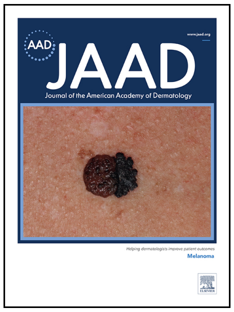


Transcriptomic analysis of cutaneous squamous cell carcinoma reveals a multi-gene prognostic signature associated with metastasis

Jun Wang, PhD, Catherine A. Harwood, MD PhD, Emma Bailey, PhD, Findlay Bewicke-Copley, PhD, Chinedu Anthony Anene, PhD, Jason Thomson, MD, Mah Jabeen Qamar, Rhiannon Laban, Craig Nourse, PhD, Christina Schoenherr, PhD, Mairi Treanor-Taylor, MD, Eugene Healy, MB PhD, Chester Lai, BM PhD, Paul Craig, Colin Moyes, William Rickaby, Joanne Martin, PhD, Charlotte Proby, MD, Gareth J. Inman, PhD, Irene M. Leigh, CBE, DSc

PII: S0190-9622(23)02504-5

DOI: https://doi.org/10.1016/j.jaad.2023.08.012

Reference: YMJD 17885

To appear in: *Journal of the American Academy of Dermatology*

Received Date: 21 April 2023

Revised Date: 26 July 2023

Accepted Date: 1 August 2023

Please cite this article as: Wang J, Harwood CA, Bailey E, Bewicke-Copley F, Anene CA, Thomson J, Qamar MJ, Laban R, Nourse C, Schoenherr C, Treanor-Taylor M, Healy E, Lai C, Craig P, Moyes C, Rickaby W, Martin J, Proby C, Inman GJ, Leigh IM, Transcriptomic analysis of cutaneous squamous cell carcinoma reveals a multi-gene prognostic signature associated with metastasis, *Journal of the American Academy of Dermatology* (2023), doi: https://doi.org/10.1016/j.jaad.2023.08.012.

Article type: Original article

Transcriptomic analysis of cutaneous squamous cell carcinoma reveals a multi-gene 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,[a,b] Rhiannon Laban,[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,[h] Colin Moyes,[i] William Rickaby,[j] Joanne Martin, PhD,[a] Charlotte Proby, MD,[k] Gareth J. Inman, PhD,[d,e] Irene M. Leigh, CBE, DSc[a]

a. Faculty of Medicine and Dentistry, Queen Mary University of London, London E1 1BB
b. Department of Dermatology, Royal London Hospital, Barts Health NHS Trust E1 IBB
c. Centre for Biomedical Science Research, School of Clinical and Applied Sciences, Leeds Beckett University, Leeds, UK
d. Cancer Research UK Beatson Institute, Bearsden Glasgow, G61 1BD
e. School of Cancer Sciences, University of Glasgow, Bearsden G61 1QH
f. Dermatopharmacology, University of Southampton, Southampton General Hospital, SO16 6YD
g. Dermatology, University Hospital Southampton NHS Foundation Trust, SO16 6YD
h. Cellular Pathology, Gloucestershire Hospitals NHS Foundation Trust, Cheltenham General Hospital, Cheltenham GL537AN
i. Queen Elizabeth University Hospital Glasgow G51 4TF
j. University College London NHS Trust, London W1T 4EU
k. Molecular and Clinical Medicine, School of Medicine, University of Dundee, DD 1 4HN

Corresponding author
Jun Wang PhD, Barts Cancer Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ. E-mail: j.a.wang@qmul.ac.uk

Word count: 2,487

Number of references: 36

Number of Figures: 3; Number of Tables: 2

Supplemental Material: https://data.mendeley.com/datasets/z77kdgddm9/1

Funding source: The research was funded by Sanofi-Regeneron as an investigator support award to IML with co-investigators JW and GJI

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

Patient consent Not applicable

ABSTRACT

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 tumours at high risk of metastasis would have a significant impact on management.

Objective: To develop a robust and validated gene expression profile (GEP) 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 non-metastasising and 86 metastasising) were collected retrospectively from four centres. 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-centre study and larger prospective multicentre studies are now required.

Conclusion: The 20-gene signature prediction is accurate, with the potential to be incorporated into clinical workflows for cSCC.

Key words:

Cutaneous squamous cell carcinoma; Metastasis; Prognosis; Transcriptomics; Machine learning; Risk stratification

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.

BACKGROUND

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

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 tumours, 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 tumour and perilesional normal skin. A range of machine learning (ML) techniques was applied and a 20-gene GEP model was developed which displayed a high level of accuracy in differentiating metastasising and non-metastasising 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 266559 (Diagnostic marker panel development for progression in skin cancer, PERMEDID). Four collaborating pathology centres identified consecutive patients with primary cSCC which had metastasised, or primary cSCC which had not metastasised within 3 years (**Table I).** Immunosuppressed patients were excluded. Formalin fixed paraffin embedded (FFPE) sections were reviewed by an expert dermatopathologist and tumour and perilesional normal skin marked for subsequent analysis (see Supplemental Materials).

Pathology review and pathological tumour staging

Haematoxylin and eosin (H&E) stained sections were digitally scanned by Leica scanner and Aperio software. Images were reviewed centrally by two expert dermatopathologists and primary tumours 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 pre-processing and normalisation 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 over-representation 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 (i.e., metastasising versus non-metastasising cSCC), the caret R package (18) was used for machine learning (ML) analysis. A range of ML techniques were used and compared (Supplemental Materials). We randomly split the samples into training (75%) and testing (25%) sets. Starting with an initial set of genes in the training set (i.e. all DE genes from the DE analysis comparing metastasising and non-metastatic cSCC), the best performing set of genes for each ML algorithm (i.e., 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, positive predive value (PPV) and negative predictive value (NPV).

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 Supplemental 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

Gene expression profiles (GEP) of 19,072 genes across a total of 433 samples were sufficiently profiled for analysis. Four sample groups were compared; cSCC tumour from metastasising (n=84) and non-metastasising (n=146) cSCC, and matched perilesional normal skin from metastasising (n=71) and non-metastasising (n=132) cSCC (Supplemental Table I). Principal component analysis based on genes across all samples showed a clear separation between cSCC and perilesional normal skin samples from metastasising and non-metastasising cSCC (Supplemental Fig 1). Differential gene expression analysis revealed that 1,038 genes were upregulated and 236 genes downregulated in metastasising cSCC compared to non-metastasising cSCC (absolute $\log_2$ fold change >1 and adjusted *p*-value <0.05). The gene set over-representation test showed keratinisation, B-cell receptor (BCR), innate immune response, cell cycle, DNA replication and DNA repair were highly over-represented in the DE genes (hypergeometric test *q*<0.05, Supplemental Fig 2A). Over-representation analysis against cellular signatures showed that signatures associated with neural progenitor, endothelial and cancer stem cells were highly enriched within the DE genes (Supplemental Fig 2B), 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 signalling), were all significantly upregulated in metastasising cSCC, while formation of the cornified envelope, keratinisation, and many metabolism pathways (sphingolipid, triglyceride, creatine and fatty acid metabolism) were significantly downregulated (**Figure 1**).

Normal perilesional samples from metastasising and non-metastasising primary cSCC were also compared. GSEA indicated many immune pathways (such as BCR and T cell receptor signalling, 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 metastasising tumours (Supplementary Table II).

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 neighbours (KNN) was identified (Supplemental Table III) which provided the best performance in differentiating the two cSCC groups in the validation set (n=57: 36 non-metastasising; 21 metastasising), 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 versus 21.7% for high-risk) (**Figure 2**). In this 20-gene GEP model, 18 genes were upregulated in non-metastasising cSCC and 2 genes (*MDK* and *STMN1)* were upregulated in metastasising cSCC (Supplemental Table III, Supplemental Fig 3). Functional annotation of the 20 genes suggested the significant enrichment in the signatures from keratinisation, GnRH, oxytocin, Ras and MAPK signalling pathways (hypergeometric test, p<0.01). 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: 49.2-77.1%), 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, tumours were staged by both UICC-8 TNM and BWH T-staging classifications after central 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**, Supplemental Table IV).

The 20-GEP signature showed strong correlations with staging for risk prediction in the validation set. Of 23 metastasising cases predicted by the 20-GEP test, 21/23 (91.3%) were T2b/T3 by BWH staging versus 15/23 (65.2%) UICC-8 T3/4. Of 32 non-metastasising 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 tumours 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 area under the ROC curve (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 (**Figure 3**). In comparison, the KNN binary classification model (i.e., 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 metastasising versus non-metastasising cSCC in both training and testing sets (Wilcoxon rank sum test, $p<0.0001$, Supplemental Fig 4).

Finally, the linear predictors across both tumour and perilesional skin for both metastasising and non-metastasising cSCC were compared (Supplemental Fig 5). There was no difference in linear predictors between non-metastasising cSCC and both normal adjacent groups. However, linear predictors increased significantly for metastasising cSCC compared to other groups ($p<0.0001$), suggesting that our linear predictor was only associated with metastasising primary tumours.

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 positive predictive value 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.

There appeared to be a strong association between the 20-GEP and keratinisation. Key keratinisation genes, such as *LCE1C*, *LCE2B/C*, *LCE3C* and *CDSN*, were all significantly downregulated in metastasising 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 *MDK*, were significantly upregulated in metastasising samples. STMN1 a microtubule-destabilising protein, regulates the dynamics of microtubules and cell cycle progress (20). Its high expression is associated with poor prognosis in oesophageal (ESCC), lung (LUSC) and oral SCC (21-23). In ESCC and LUSC, 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). Midkine (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 tumour 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 non-standardised reporting of high-risk features (particularly poor differentiation and perineural invasion); problems defining the state of differentiation of an individual tumour; 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 on 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 centre, 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.

Abbreviations used:

AJCC: American Joint Committee on Cancer

UICC: Union for International Cancer Control

BWH: Brigham and Women’s Hospital

cSCC: cutaneous squamous cell carcinoma

GEP: gene-expression profile

DE: differential expression

GSEA: Gene set enrichment analysis

HR: hazard ratio

LR: likelihood ratio

NPV: negative predictive value

PPV: positive predictive value

KNN: K-nearest neighbourhood

BCR: B-cell receptor

Author contributions

IML, JW and GJI conceived and designed the study, and acquired the funding. CAH, CP, EH, CL, CM, JT and MJQ recruited the patient cohort and collected the clinical data. PC, WR and JM performed the tumour grading and histology analysis. EB, JW, FBC and CAA performed the bioinformatics and machine learning data analysis. CN, CS and MTT performed the TempO-seq experiment. JW, IML, GJI and CAH supervised the study, analysed and interpreted the data, and drafted the manuscript. All authors critically revised the manuscript and approved the final version to be submitted.

Acknowledgements

JW, EB, FBC and CAA acknowledge support from Cancer Research UK City of London Major Centre core funding support to Barts Cancer Institute (C16420/A18066). JW, CH, EB and IML also acknowledge funding support from The Barts Charity Strategic Award of Barts Centre and Squamous Cancer (MGU0603). JW also acknowledges support from Cancer Research UK (C355/A26819) and FC AECC and AIRC under the Accelerator Award Program, and support from the Academy of Medical Sciences Springboard Award (SBF003\1025). CN, CS, MTT and GJI acknowledge support from Cancer Research UK (A29802) and MTT is funded by a Cancer Research UK-Funded TRACC PhD studentship. CL is funded by the MRC Clinical Academic Research Partnerships (CARP).

References

1. Venables ZC, Nijsten T, Wong KF, Autier P, Broggio J, Deas A, Harwood CA, Hollestein LM, Langan SM, Morgan E, Proby CM, Rashbass J, Leigh IM. Epidemiology of basal and cutaneous squamous cell carcinoma in the U.K. 2013-15: a cohort study. Br J Dermatol. 2019 Sep;181(3):474-482.

2. Venables ZC, Autier P, Nijsten T, Wong KF, Langan SM, Rous B, Broggio J, Harwood C, Henson K, Proby CM, Rashbass J, Leigh IM. Nationwide Incidence of Metastatic Cutaneous Squamous Cell Carcinoma in England. JAMA Dermatol. 2019 Mar 1;155(3):298-306.

3. Kwiatkowska M, Ahmed S, Ardern-Jones MR, Bhatti LA, Bleiker TO, Gavin A, Hussain S, Huws DW, Irvine L, Langan SM, Millington GWM, Mitchell H, Murphy R, Paley L, Proby CM, Thomson CS, Thomas R, Turner C, Vernon S, Venables ZC. A summary of the updated report on the incidence and epidemiological trends of keratinocyte cancers in the UK 2013-2018. Br J Dermatol. 2022 Feb;186(2):367-369.

4. Detailed statistics from the Get Data Out (>Skin Tumours) programme, National Disease Registration Service, UK. https://www.cancerdata.nhs.uk/getdataout/skin

5. Venables ZC, Tokez S, Hollestein LM, Mooyaart AL, van den Bos RR, Rous B, Leigh IM, Nijsten T, Wakkee M. Validation of four cutaneous squamous cell carcinoma staging systems using nationwide data. Br J Dermatol. 2022 May;186(5):835-842.

6. Tokez S, Venables ZC, Hollestein LM, Qi H, Bramer EM, Rentroia-Pacheco B, van den Bos RR, Rous B, Leigh IM, Nijsten T, Mooyaart AL, Wakkee M. Risk factors for metastatic cutaneous squamous cell carcinoma: Refinement and replication based on 2 nationwide nested case-control studies. J Am Acad Dermatol. 2022 Jul;87(1):64-71.

7. South AP, Purdie KJ, Watt SA, Haldenby S, denBreem N, Dimon M, Arron ST, McHugh A, Xue DJ, Jasbani HS, Dayal HS, Proby CM, Harwood CA, Leigh IM, Prevalent NOTCH1 mutations in squamous cell carcinogenesis are an early event J Invest Dermatol 2014 :134; 2630-8

8. Cammameri P, Rose AM, Vincent DF, Wang J, Nagano A, Libertini S, Ridgeqay RA, Athineos D, Coates PJ, McHugh A, Pourreyron C, Dayal JH, Larsson J, Weidlich S, Spender LC, Sapkota GP, Purdie KJ, Proby CM, Harwood CA, Leigh IM, Clevers H, Barker N, Karlsson S, Pritchard C, Marais R, Chelala C, South AP, Sansom OJ, Inman GJ. Inactivation of TGFb receptors in stem cell drives cutaneous squamous cell carcinoma. Nature Commun 2016 25: 12493

9. Inman GJ, Wang AJ, Ai N, Alexandreev L, Chelala C, Stratton M, Harwood CA, Sherwood V, Proby CM, Leigh IM. 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 Sept 10 9(1) 3667.

10. Thomson J, Bewicke-Copley F, Anene CA, Gulati A, Nagano A, Purdie K, Inman GJ, Proby CM, Leigh IM, Harwood CA,Wang J. The genomic landscape of actinic keratosis. J Invest Dermatol 2021 Jul;141(7):1664-1674.e7.

11. Wysong A, Newman JG, Covington KR, Kurley SJ, Ibrahim SF, Farberg AS, Bar A, Cleaver NJ, Somani AK, Panther D, Brodland DG, Zitelli J, Toyohara J, Maher IA, Xia Y, Bibee K, Griego R, Rigel DS, Meldi Plasseraud K, Estrada S, Sholl LM, Johnson C, Cook RW, Schmults CD, Arron ST. 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 Feb;84(2):361-369.

12. Ibrahim SF, Kasprzak JM, Hall MA, Fitzgerald AL, Siegel JJ, Kurley SJ, Covington KR, Goldberg MS, Farberg AS, Trotter SC, Reed K, Brodland DG, Koyfman SA, Somani AK, Arron ST, Wysong A. Enhanced metastatic risk assessment in cutaneous squamous cell carcinoma with the 40-gene expression profile test. Future Oncol. 2022 Mar;18(7):833-847. doi: 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, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

15. Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Zhang Y, Storey JD, Torres LC. Sva: Surrogate Variable Analysis. 2022 R package version 3.44.0.

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, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021 Jul 1;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, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17;12:77.

20. Rubin CI, Atweh GF. The role of stathmin in the regulation of the cell cycle. J Cell Biochem. 2004 Oct 1;93(2):242-50.

21. Suzuki S, Yokobori T, Altan B, Hara K, Ozawa D, Tanaka N, Sakai M, Sano A, Sohda M, Bao H, Fukuchi M, Miyazaki T, Kaira K, Asao T, Kuwano H. High stathmin 1 expression is associated

with poor prognosis and chemoradiation resistance in esophageal squamous cell carcinoma. Int J Oncol. 2017 Apr;50(4):1184-1190.

22. Bao P, Yokobori T, Altan B, Iijima M, Azuma Y, Onozato R, Yajima T, Watanabe A, Mogi A, Shimizu K, Nagashima T, Ohtaki Y, Obayashi K, Nakazawa S, Bai T, Kawabata-Iwakawa R, Asao T, Kaira K, Nishiyama M, Kuwano H. High STMN1 Expression is Associated with Cancer Progression and Chemo-Resistance in Lung Squamous Cell Carcinoma. Ann Surg Oncol. 2017 Dec;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 Oct;44(10):1725-1732.

24. Ni PZ, He JZ, Wu ZY, Ji X, Chen LQ, Xu XE, Liao LD, Wu JY, Li EM, Xu LY. Overexpression of Stathmin 1 correlates with poor prognosis and promotes cell migration and proliferation in oesophageal squamous cell carcinoma. Oncol Rep. 2017 Dec;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 Feb;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 Feb;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 Mar;39(10):2040-2054.

28. Yu X, Zhou Z, Tang S, Zhang K, Peng X, Zhou P, Zhang M, Shen L, Yang L. MDK induces temozolomide resistance in glioblastoma by promoting cancer stem-like properties. Am J Cancer Res. 2022 Oct 15;12(10):4825-4839.

29. Olmeda D, Cerezo-Wallis D, Riveiro-Falkenbach E, Pennacchi PC, Contreras-Alcalde M, Ibarz N, Cifdaloz M, Catena X, Calvo TG, Cañón E, Alonso-Curbelo D, Suarez J, Osterloh L, Graña O, Mulero F, Megías D, Cañamero M, Martínez-Torrecuadrada JL, Mondal C, Di Martino J, Lora D, Martinez-Corral I, Bravo-Cordero JJ, Muñoz J, Puig S, Ortiz-Romero P, Rodriguez-Peralto JL, Ortega S, Soengas MS. Whole-body imaging of lymphovascular niches identifies pre-metastatic roles of midkine. Nature. 2017 Jun 28;546(7660):676-680.

30. Cerezo-Wallis D, Contreras-Alcalde M, Troulé K, Catena X, Mucientes C, Calvo TG, Cañón E, Tejedo C, Pennacchi PC, Hogan S, Kölblinger P, Tejero H, Chen AX, Ibarz N, Graña-Castro O,

Martinez L, Muñoz J, Ortiz-Romero P, Rodriguez-Peralto JL, Gómez-López G, Al-Shahrour F, Rabadán R, Levesque MP, Olmeda D, Soengas MS. Midkine rewires the melanoma microenvironment toward a tolerogenic and immune-resistant state. Nat Med. 2020 Dec;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 Aug 26;13(1):5018.

32. Jambusaria-Pahlajani A, Kanetsky PA, Karia PS, Hwang WT, Gelfand JM, Whalen FM, Elenitsas R, Xu X, Schmults CD. Evaluation of AJCC tumor staging for cutaneous squamous cell carcinoma and a proposed alternative tumor staging system. JAMA Dermatol. 2013 Apr;149(4):402-10. doi: 10.1001/jamadermatol.2013.2456. PMID: 23325457.

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. doi: 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. doi:10.1001/jamadermatol.2019.0032

36. Blamey RW, Ellis IO, Pinder SE, Lee AH, Macmillan RD, Morgan DA 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.

Figures Legend

Fig 1. Normalised enrichment scores (NES) of the top dysregulated canonical pathways between metastasising and non-metastasising cSCC. Pathways with positive NES (in red) were upregulated while pathways with negative NES (in blue) were downregulated in metastasising compared to non-metastasising primary cSCC.

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.

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
Tumour diameter, cm, mean (range) #	1.85 (0.18-9)	1.31 (0.18-4.1)	2.82 (1.6-9)	<.0001
Tumour 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 (≥ 0.1mm)	20 (8.6)	8 (5.3)	12 (14.6)	
Present (<0.1mm 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 metastasise =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 non-metastasising)

# not available for 10 cSCC (5 non-metastasising and 5 metastasising)

## not available for 15 cSCC (10 non-metastasising and 5 metastasising)

§ Invasion through or beyond subcutaneous fat: not available for 7 cSCC (3 non-metastasising and 4 metastasising)

¶ not available for 5 cSCC (1 non-metastasising and 4 metastasising cSCC)

∞ Lymphovascular invasion not available for 6 cSCC (all metastasising)

§§ not available for 12 cSCC (6 non- metastasising and 6 metastasising)

¶¶ not available for 10 cSCC (4 non-metastasising and 6 metastasising)

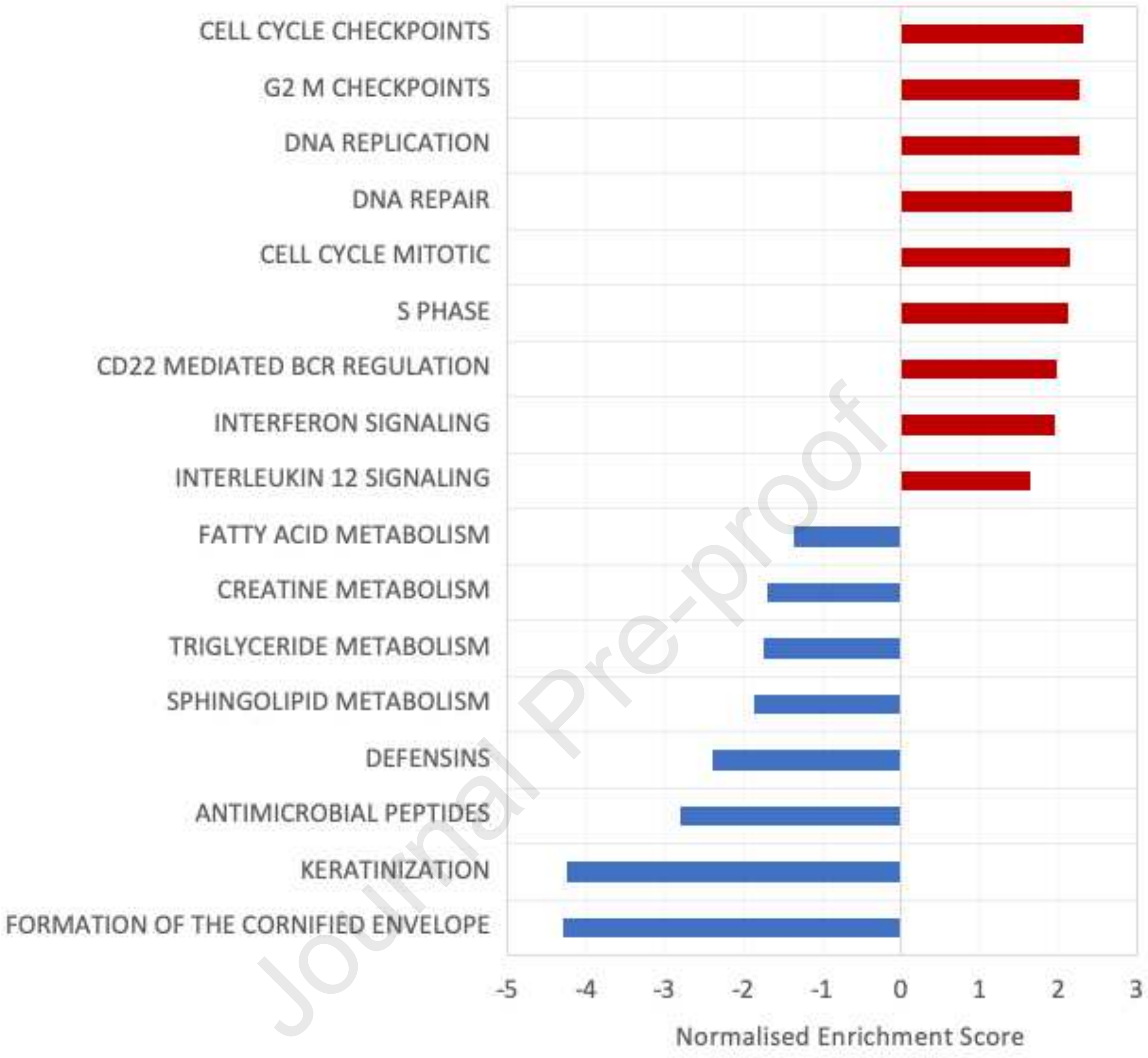
Table II. Accuracy of the prediction of metastatic risks of the 20-GEP signature and other risk assessment methods (n=57).

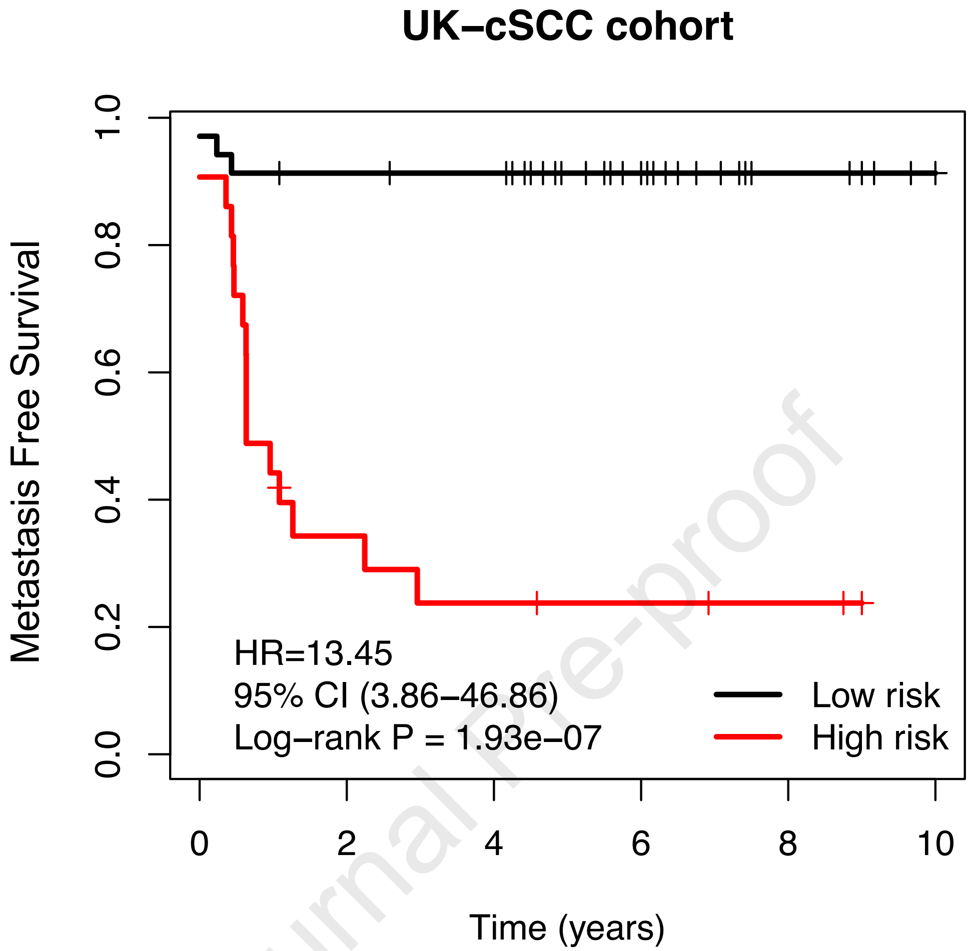
Classifier	Accuracy%	Sensitivity%	Specificity%	PPV%	NPV%	+LR	-LR
20-GEP	86.0	85.7	86.1	78.3	91.2	6.17	0.17
UICC-8	85.4	81.0	88.2	81.0	88.2	6.88	0.22
BWH	85.4	95.2	79.4	74.1	96.4	4.63	0.06
BWH v1	81.8	76.2	85.3	76.2	85.3	5.18	0.28
22-GEP*	64.0	41.2	75.8	46.7	71.4	1.70	0.78
UICC-8**	76.5	58.8	86.3	70.1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

UICC: Union for International Cancer Control; 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; PPV: Positive Predictive Value; NPV: Negative Predictive Value; +LR: Positive Likelihood Ratio; -LR: Negative Likelihood Ratio.

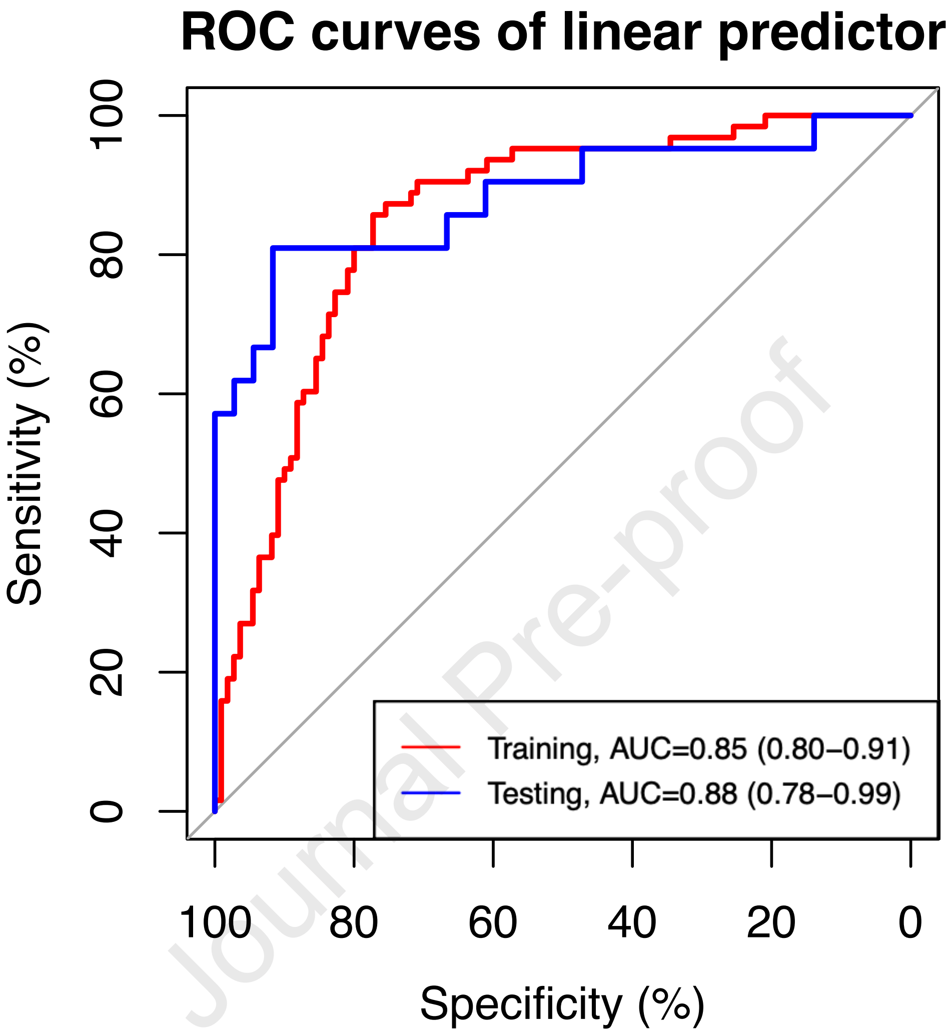
22-GEP* was derived from normal adjacent samples only.

** Statistics were derived from the whole cohort (n=237)


CELL CYCLE CHECKPOINTS
G2 M CHECKPOINTS
DNA REPLICATION
DNA REPAIR
CELL CYCLE MITOTIC
S PHASE
CD22 MEDIATED BCR REGULATION
INTERFERON SIGNALING
INTERLEUKIN 12 SIGNALING
FATTY ACID METABOLISM
CREATINE METABOLISM
TRIGLYCERIDE METABOLISM
SPHINGOLIPID METABOLISM
DEFENSINS
ANTIMICROBIAL PEPTIDES
KERATINIZATION
FORMATION OF THE CORNIFIED ENVELOPE
-5
-4
-3
-2
-1
0
1
2
3
Normalised Enrichment Score



# at risk	0	2	4	6	8	10
Low-risk	34	30	29	17	6	2
High-risk	21	6	4	3	2	0


ROC curves of linear predictor
Sensitivity (%)
0
20
40
60
80
100
Specificity (%)
100
80
60
40
20
0
Training, AUC=0.85 (0.80–0.91)
Testing, AUC=0.88 (0.78–0.99)