

# Whole Transcriptomic Analysis of Patient Derived Skin Lesions Reveals Key Driver Mechanisms of Skin Squamous Cell Carcinoma

Xiu Cheng Quek<sup>1</sup> Miko Yamada<sup>2</sup>, Konstantin Shakhbazov<sup>3</sup>, Lynlee L. Lin<sup>2</sup>, Elizabeth J. Payne<sup>2</sup>, Mark J. Cowley<sup>1</sup>, Sudipta Sinnya<sup>2</sup>, Jean M. Tan<sup>2</sup>, Ian H. Frazer<sup>3</sup>, H. Peter Soyer<sup>2</sup>, Tarl W. Prow<sup>2</sup>, Marcel E. Dinger<sup>1</sup>

<sup>1</sup> Garvan Institute of Medical Research, Sydney, Australia,

<sup>2</sup> Dermatology Research Centre, The University of Queensland, School of Medicine, Transitional Research Institute at Princess Alexandra Hospital, Woolloongabba, Brisbane, Australia 4102,

<sup>3</sup> Diamantina Institute, The University of Queensland, Transitional Research Institute at Princess Alexandra Hospital, Brisbane, Australia, 4021

## Introduction

Skin Squamous Cell carcinoma (SCC) is cancer results from the uncontrolled proliferation of epithelial cells and is the second most common form of cancer that in Australia.

97% of invasive SCC (iSCC) develops from actinic keratosis (AK), a skin condition characterized by loss of the granular layer and thickening of the epidermis. However, histological distinction between AK, iSCC and an intermittent stage, referred as Intraepidermal Squamous Cell carcinoma (IEC), is difficult and requires expert knowledge. In addition, treatments for AK, IEC and iSCC varies and hence the study of AK progression to SCC has been for key importance.



Figure 1 : Progression of iSCC from AK is marked by thickening of the epidermis layer (green) and loss of granular layer (red).

## Aim

To further investigate the biology of AK progression to iSCC, we conducted the first ever study of SCC that utilizes RNA-seq to investigate the transcriptomic landscape of SCC progression.

Using data from the transcriptome, we aim to discover the following:

- 1) Whole transcriptome identification of differentially expressed genes (DGE) between the three conditions.
- 2) Elucidate a possible role of long-non-coding RNAs in the disease.
- 3) From the DGE, determine the pathway and mechanism driving the different stages of disease
- 4) Identify gene candidates for therapeutic purposes.

## Methods

15 AK, 8 SCC and 8 IEC skin lesion from a cohort of 25 patients was obtained. was isolated and subjected to RNAseq using the HiSeq Illumina platform

RNA-Seq reads were aligned using Tophat2 and gene expression summarized based on the GENCODE Ver17 gene models.

Differentially expressed genes (DGE) was conducted using edgeR.

Functional enrichment analysis was conducted using the Gene set enrichment tool kit.

RNA variant calling conducted with Samtools and variants characterized with PROVEAN

## Results

A total of ~600 to 1000 genes are misregulated between the three conditions.

Misregulated Genes Between Pairwise Comparison of AK, IEC and iSCC

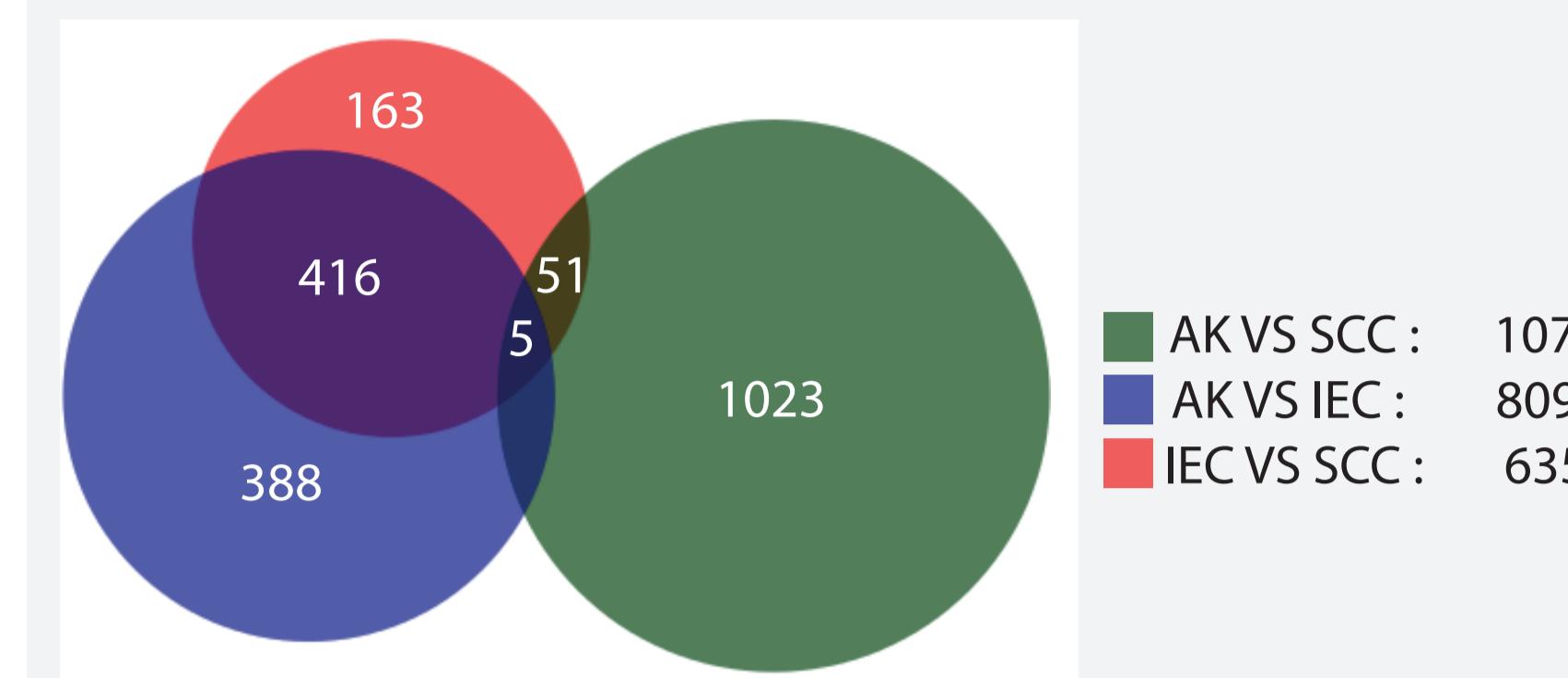


Figure 2: Comparision of gene expression between AK and iSCC show the highest number of misregulated genes (green)

Long non coding RNAs are differentially expressed between AK, IEC and iSCC

We observed ~100 to ~150 lncRNAs that are differentially expressed in the three conditions. From which, we selected a number of lncrna candidates based on expression profile for further investigation (Figure 3).

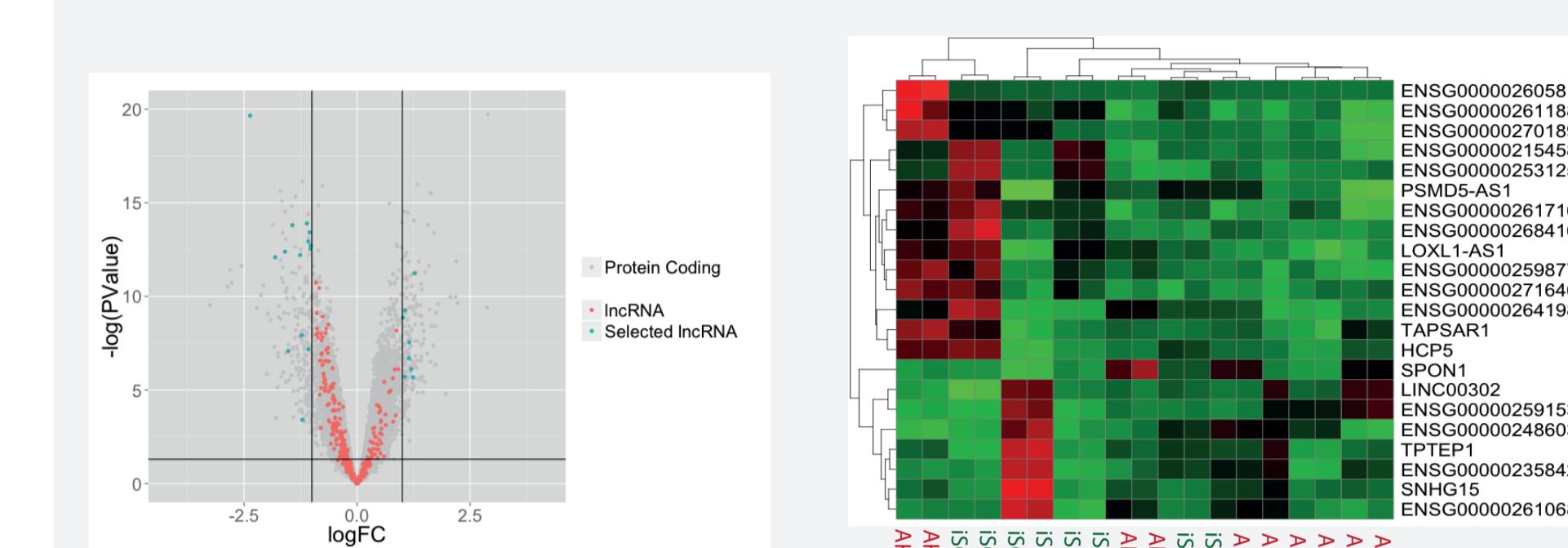
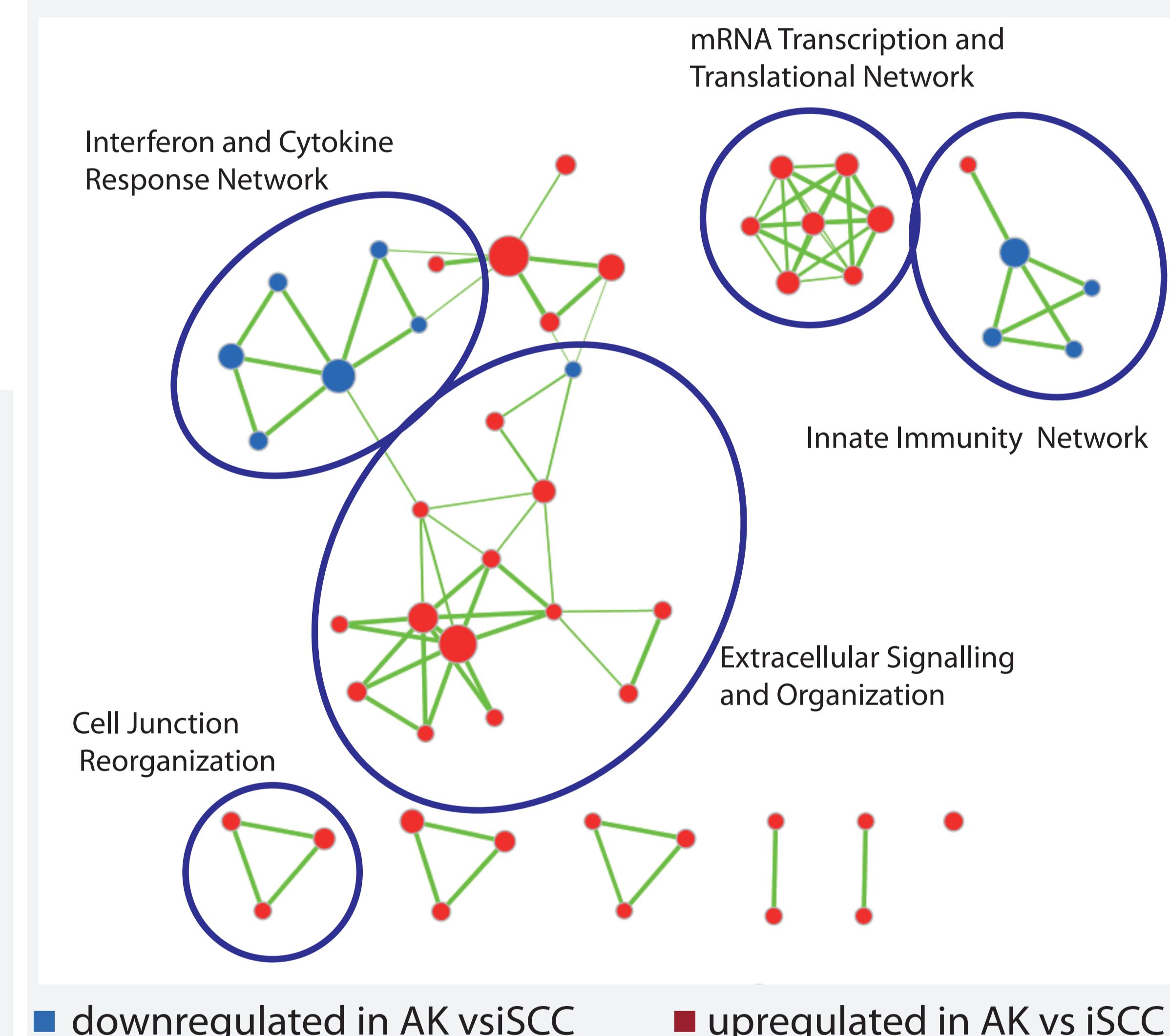


Figure 3: Selection lncrna candidates using comparison between AK and iSCC as an example. Left: Volcano plot used to select lncrna candidates ( $\log_2 FC > 1$  and  $P\text{-Value} < 0.05$ ). A total of 22 candidates were selected. Right: Heatmap showing expression of the 22 candidates

GSEA identify key biological features involved in pathogenesis of SCC

Using data obtained from comparision between AK and iSCC as an example, results from pre-ranked GSEA using log fold change value obtained from edgeR has allowed us to conduct the following:

- A) Identification of common biological themes involved in the pathogenesis of SCC (Figure 4A).
- B) From the Extracellular Signalling and Organization, we selected the Reactome extracellular matrix organization pathway (ECM) as a pathway of interest which have direct biological role in the disease (Figure 4B),
- C) Identification of specific gene candidates, including members of the ECM, for futher investigating (Figure 4C)



■ downregulated in AK vs iSCC ■ upregulated in AK vs iSCC

Figure 4A: Common biological themes (blue circle) among gene-set misregulated between AK and iSCC were identified using enrichment map analysis. Enrichment map analysis clusters mutually overlapping gene-sets (red/blue dots) from GSEA together into a network. Enrichment map analysis is conducted using the enrichment map cytoscape plugin.

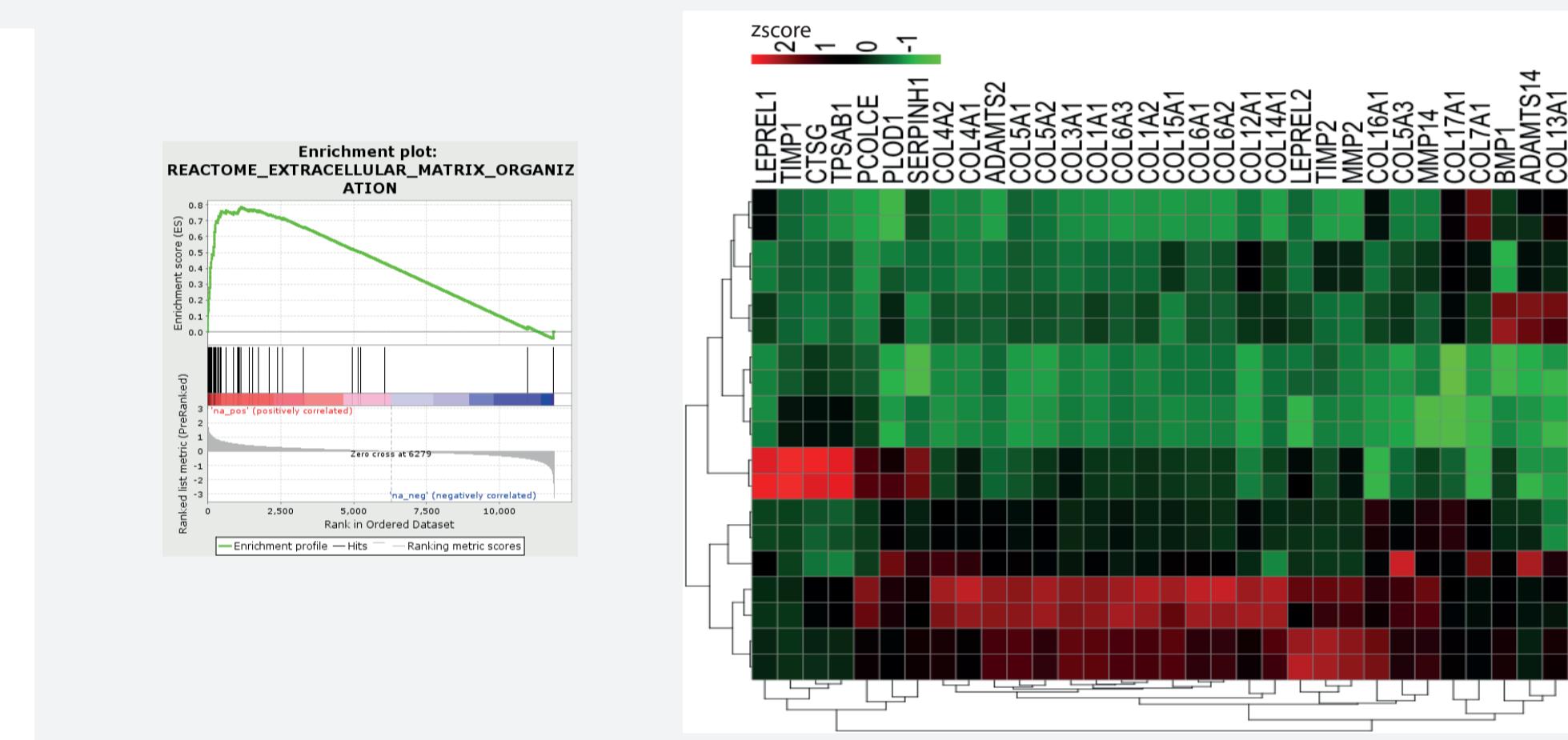


Figure 4B: The ECM pathway is upregulated in AK verus iSCC( $FDR < 0.0001$ ). Left: Snapshot of GSEA enrichment results for the ECM pathway . Right: Heatmap of genes involved in the ECM that are differentially expressed between AK and iSCC (  $P\text{ Value} < 0.05$  ).

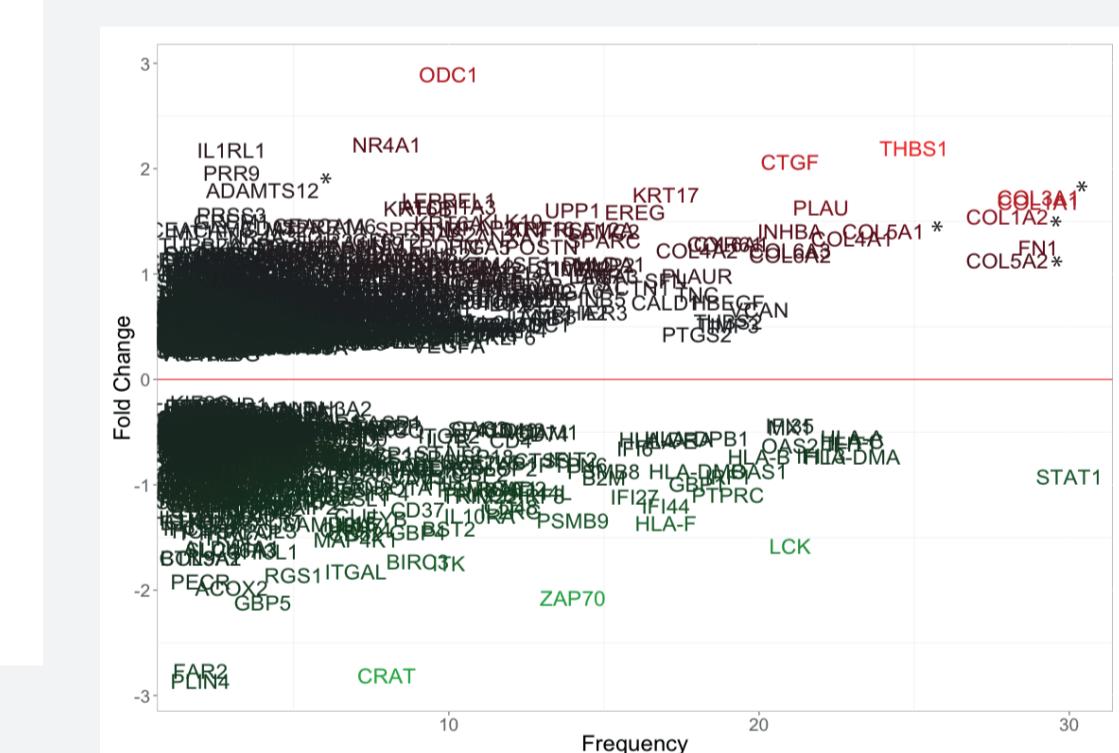
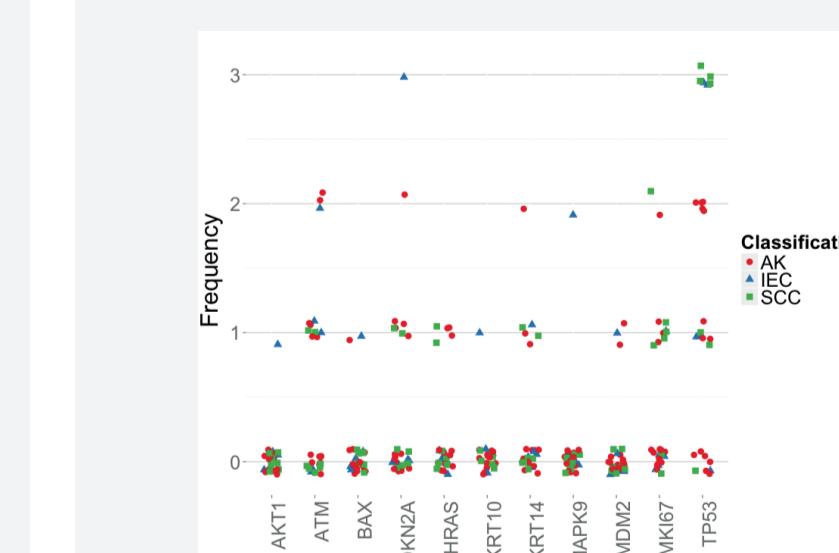


Figure 4C: Identification of specific gene candidates. Meta-GSEA displaying the frequency of an given gene in the top 100 enriched gene sets against the gene's expression value (log2fold change). Members of the EMC (asterisk) occur frequently in various genes sets and are also highly expressed. Meta-GSEA also allow identification of other gene candidates of investigation

Variant calling with RNAseq detects deleterious mutation in cancer associated genes



To further select gene candidates for investigation, using data from our RNAseq experiment, we conducted variant calling on a selected panel of 11 genes commonly associated with non-melanoma skin cancer.

Figure 5: Increase frequency of SNP mutation in TP53, MKI67 and ATM detected in iSCC samples from RNAseq variant calling. Deleterious snps were detected by PROVEAN (-2.5 threshold).

## Conclusion

- 1) DGE analysis displayed significant differences in expression AK, IEC and iSCC samples.
- 2) Misregulation of lncrna in AK, IEC and iSCC suggest a role of lncrna in SCC pathogenesis
- 3) GSEA analysis showed enrichment for various biological themes that are distinct for each of the three conditions.
- 4) For instance, the ECM pathway, belonging to the gene-set cluster involved in extracellular signalling and Organization is provided as an example whose activity is shown to be enhanced in AK when compared to iSCC.
- 5) Genes of the ECM pathway may be pivotal to the development of iSCC from AK.
- 6) Variant calling identified TP53, MKI67 and ATM to have an increased frequency of deleterious SNP mutation in iSCC