UNIVERSITY OF CALIFORNIA
SANTA CRUZ

A framework for analyzing wound transcriptome data of skin and oral mucosa

A thesis submitted in partial satisfaction
of the requirements for the degree of
MASTER OF SCIENCE
in
SCIENTIFIC COMPUTING AND APPLIED MATHEMATICS
by
Krishnakant Dasika
June 2020

The thesis of
Krishnakant Dasika is approved:

_____
Professor Qi Gong (Chair)

_____
Professor Hongyun Wang

_____
Assistant Professor Marcella M. Gomez

_____
Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Copyright © by

Krishnakant Dasika

2020

Table of Contents

List of Figures

List of Tables

Abstract

A framework for analyzing wound transcriptome data of skin and oral mucosa
by Krishnakant Dasika

Analysis of gene transcriptome data plays a crucial role in understanding key biological processes that govern wound healing. Oral mucosal wounds are found to heal much faster than skin wounds. With the aim of improving rates of wound closure in skin tissues, a comparative study of the two tissues in different wound healing stages is undertaken. Gene transcriptome data of skin and tongue tissues in *mus musculus* is used to develop a general framework for analyzing gene microarray data and extract meaningful observations. The author presents a new approach to clustering gene time-series dynamics, taking the underlying biological processes into consideration. Furthermore, an analysis on the comparison of skin and tongue tissues' healing is presented, highlighting processes that are unique to the two tissues based solely on clustering gene microarray expression data.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor Dr. Marcella M. Gomez for her mentorship and support throughout the project. This work could not have been completed without her guidance and motivation. I would also like to thank my team members Ksenia Zlobnia, Jianhong Chen, Mohammad Jafari and Giovanny Marquez for their invaluable inputs and feedback. Finally I would like to thank all the faculty of the Applied Mathematics department for giving me this wonderful opportunity to pursue my masters at the University of California, Santa Cruz.

1.1 Introduction

A thorough understanding of molecular and systemic functions during different stages of wound healing potentially helps drive research into the development of novel medical devices targeted at accelerating the rate of wound closure. One such method is the design of smart bandages which improve macrophage functionality using bioelectric signals to the surrounding tissues.

Some recent studies are devoted to bandages for control and treatment of wound healing [18] [19]. Understanding of wound healing stages on the level of gene expression may help in creating bioelectronics wound healing control systems [20]. Studies [21] also suggest that the use of bio-electrical mechanisms may enhance the rate of wound closure by improved recruitment rates of epithelial and connective tissue cells. Therefore, for proper intervention techniques, the dynamics of chemokines, cytokines, interleukins and keratins secreted by cells participating in inflammatory response and proliferation need to be analyzed in detail. Since wound healing in tongue tissue is established to be faster than in skin tissue [17], a comparison of the processes between the two tissues could help identify the pathways that need to be manipulated to achieve higher wound closure rates.

Chen et al [2] have proposed a systematic framework distinguish the differences in gene expression between skin and tongue tissue of mus musculus. They employ the use of Affymetrix cDNA microarray data which is divided into groups based on k-means clustering, followed by gene annotations to confirm the difference between the two tissues. The framework helps identify crudely the gene groups in the two tissues and present a case for stronger inflammatory response in skin tissue compared to tongue tissue. J.Cheng [13] extended this work by presenting a Boolean network approach to identify the status of the wound. Using an example of six selected functional groups, the author presents a method to identify the stage of wound healing based on whether the functional group is active or inactive at the time of biopsy. It should be noted that both investigations use data clustering in general as the mathematical method for gene expression analysis. Upon investigating the methods of clustering used, it was found that using the entire data set, which usually consist of tens of thousands of gene dynamics, led to sub-optimal clustering results. The obtained gene groups could only be considered as crude estimates for the gene functionality

at each stage. This called for a need to improve clustering results. Most studies on gene expression data categorize the results into broadly two groups – early and late expressing genes. In order to pinpoint the differences in dynamics between the two tissues, there is a need to identify intermediate gene clusters as well. In addition to the result presented in Chen et al, , we would like to identify which processes cause delayed response in skin and which processes help accelerate wound healing in tongue tissues. In the current work, we look for automatic and semi-automatic methods to find differences in gene dynamics. A new method for the analysis gene microarray data is proposed, with improved clustering results as well as a method to identify gene expression specific to skin tissues.

1.2 Overview of wound healing

The overall process of wound healing consists of four highly overlapping stages, hemostasis, inflammation, proliferation/repair, and remodeling [9]. Hemostasis is the first stage of the wound healing process and the shortest of the phases lasting from 0 to a few hours post wounding. Platelets trigger a short phase of vasoconstriction to reduce blood loss followed by clot formation which is composed primarily of fibrin-polymers and blood cells. Platelets also help trigger the infiltration of leukocytes by releasing chemotactic factors. This process paves the way for inflammatory response.

Inflammation starts within 24 hours, and goes on for 2 weeks or more. Naïve lymphocytes, B and T cells are recruited to the site of injury. Circulating the blood vessels, there are erythrocytes (red blood cells), monocytes, neutrophils (phagocytes) and inflammatory mediators. There are two types of inflammatory mediators: Plasma inflammatory mediators which include complementary proteins and kinins and cell derived inflammatory mediators which include mast cells and dendritic cells (langerhan cells in skin tissue). PAMPs expressed by pathogens infiltrating the site of injury are recognized by the immune cells present in the tissues [24]. Mast cells upon identifying the pathogen secrete histamine, which when secreted causes vasodilation and causing increased permeability of blood vessels. This process allows more inflammatory response mediators and immune cells to get transported to the site of injury. Tissue macrophages on identifying the pathogen secrete cytokines primarily TNF-alpha and IL1. Cytokines

cause two types of effects: Local effect and systemic effect. Local effect is increased inflammation by vasodilation and increased vascular permeability. The monocytes recruited from the blood vessels differentiate into macrophages. In addition, the local effect of cytokines is the triggering of tissue repair by stimulating fibroblast activity. The systemic effect of cytokines includes the development of fever and leukocytosis [23]. Complement proteins, i.e., the plasma derived inflammatory mediators with antibodies can cause optimization or lysis of pathogens. C3a and C5a are the most notable complimentary proteins as they promote inflammation, vascular permeability and also act as chemo attractants, stimulating repair.

The processes described above are part of the innate immune response. However, if a pathogen causes an infection in the body, the adaptive immune response is activated. Antigen presenting cells (APCs) such as dendritic cells or macrophages engulf the pathogen which activates the APCs causing them to travel to the lymph nodes via the lymph vessels, triggering the activation of naïve B and T cells in the lymph nodes. The naïve B cells become plasma cells which can secrete antibodies such as Ig-m, Ig-g and Ig-d and the naïve T cells become either T-helper cells or T-killer cells which then remove the pathogens and bring the system back to its normal state.

Proliferation phase essentially begins when cytokines and platelet derived growth factors (PGDF) secreted by macrophages and platelets, initiate the migration or stimulation of fibroblasts into the wound. The process of proliferation is initiated by the production of matrix metalloproteinases (MMPs) which are produced by fibroblasts. MMPs help in the fluid transport of fibroblasts within the matrix. They also play an important role in wound re-epithelialization by regulating extracellular matrix degradation and deposition [16]. Fibroblasts begin proliferating and increase the synthesis of collagen which is the first step in tissue repair. This process is aided by growth factors TGF-beta and connective tissue growth factors (CTGF). PGDF also stimulates chemotaxis which further promotes tissue repair. If there are damaged blood vessels, coagulation cascade is performed. Keratinocyte differentiation is another important aspect of this stage of wound healing. Keratin growth factors (KGF) are found to play a significant role [14] in epithelialization through stimulating epidermal and dermal regeneration. In addition, KGFs induce TGF-alpha expression and EGF-receptor signaling in tissues subject to exposure

from injuries, which partake in the mechanisms of epithelial cell proliferation leading to the reconstitution of an intact epidermis [15].

1.3 DNA Microarray Data

With a wide range of microarray experiment data available, it is essential to choose relevant cDNA data that could help distinguish molecular functions pathways between the two tissues. The main criteria used to select the data sets used in this study are: A) The data set must contain gene intensities for control i.e., unwounded tissues to identify the baseline gene expression values for the two tissues. B) The span of the experiment must include the entire duration of wound healing process, typically 0-15 days for incisional wounds. C) The intervals for biopsies should roughly include all the stages of wound healing. D) Measurements should be consistent across replicates.

The dataset available in the GEO accession GSE23006, titled "Transcriptional profiling of a wound healing process in skin and oral mucosa" was found to meet most of the above criteria and is therefore used as the primary data source for designing the framework proposed in this study. A total of six samples, three for skin tissue and three for the tongues, were measured at eight different time points covering the whole time span of the murine wound healing. For each chip, the intensity level of over 45,000 genes are measured as the raw data.

2 Methods

2.1 Normalization

The intensity values are normalized to facilitate comparison across different genes. For each gene, the maximum gene expression value is found over all 8 time points across the three. The normalized intensity $\tilde{I}(t)$ at time t is defined as:

$$\tilde{I}(t) = \frac{I(t) - I(0)}{\max_t \left(\max_i I_i(t) \right)}$$

where $I_i(t)$ is intensity in replicate i at time t and $I(t) = \frac{1}{3}\sum_i I(t)$ – the average by 3 replicates intensity.

2.2 Filtering

To gain insights into the biological processes at different stages of wound healing, we first filter the genes into 6 groups as follows:

- Set a threshold for normalized intensity value beyond which a gene is considered up-regulated or down-regulated. For the purpose of this study, the threshold is assumed to be ± 0.3 above/below the baseline value
- Three intervals are considered: 0-24 hrs, 24-120 hrs and 120-240 hrs corresponding to early, middle and late stages of wound healing
- The gene is considered to be up/down regulated in the time interval if its normalized intensity $\tilde{I}(t)$ reaches the threshold value 0.3/-0.3 for the first time at that time interval.
- Once a gene is filtered into a category, it is removed from the pool of available genes
- Downregulated genes are filtered out first followed by upregulated genes. Thus, if some gene satisfies both down and up regulated conditions then it is filtered as down regulated. The genes are filtered into 6 groups:

  Downregulated 0-24hrs, Downregulated 24-120hrs, Downregulated 120-240 hrs, Upregulated 0-24hrs, Upregulated 24-120hrs and Upregulated 120-240hrs.
- If the threshold value of 0.3 is met at the boundary of an interval, the gene is still filtered into that interval.

2.3 Clustering

For each of the 6 groups, we employ k-means clustering to identify genes having similar time-series dynamics. The motivation behind performing clustering is to decompose a large amount of genes that are mixed in terms of expression level into smaller groups based on their relative expression level over the span of the entire wound healing process. Moreover, it is relatively difficult to understand the

contribution toward the wound healing processes of each individual gene in comparison to a collective group of genes, because biological processes usually require large amounts of genes to work together in order to achieve their functionality. Naturally, clustering becomes the method of choice to achieve this objective as part of the analysis. MATLAB's k-means function which utilizes k-means++ algorithm is used, which is a data-partitioning algorithm that assigns *n* observations to exactly one of *k* clusters defined by centroids by minimizing the squared euclidean distance between observations. For this purpose, 'dist' metric was set to 'sqeucliden'. Alternatively, the 'dist' metric can also be set to 'correlation' where in the centroids are the component wise means of the points in the cluster. However, using this metric causes further normalization of values to zero mean and zero standard deviation which was not necessary. Various k values are explored and k=4 is chosen as it showed good results.

2.4 Identifying inactive gene sets

In addition to clustering, another approach is explored in this study which is used to filter genes into an 'inactive gene' category. In the comparison of gene expression between skin and tongue tissues, we find that some genes that show differential expression in skin do not show any deviation from their baseline values in tongue tissue. Identifying these genes is important as they help determine key differences in the responses to injury between the two tissues. For this purpose, we use the normalized intensity values and define an inactive gene as one having expression level within ±0.05 across all time points. In other words, a gene is inactive if:

$$-0.05 < \tilde{I}(t) < 0.05 \; \forall t \qquad (*)$$

First, a set of inactive genes in tongue is obtained by using the above filter for all the normalized gene intensity values from the entire dataset. Normalized intensity values of these filtered set of genes is observed in skin tissue and genes that do not satisfy condition (*) are then identified and classified as 'active genes' in skin. This gives us the set of genes that are active in skin tissue which are inactive in

tongue tissue. The same procedure is then carried vice versa and the set of genes active in tongue tissue but inactive in skin tissue is obtained.

3 Gene Ontology

The online gene annotation consortium called the Gene Ontology (GO) is a hierarchically organized collection of functional gene sets. It is a useful tool to define the collective biological functionality of the genes group. GO provides a large, up-to-date, and comprehensive computational model that maps biological systems from molecular levels to organism levels to biological functions. All GO annotations are ultimately supported by scientific literature either directly or indirectly. In GO, the supporting evidence is presented in the form of either a published reference or description of the methodology used to create the annotation [1][3].

In this work, genes from a cluster/group are fed into the GO analysis algorithms by first cross referencing to existing online GO databases that pull from all available sources. Web-based Gene Set Analysis Toolkit (WebGestalt) [11] is used with the reference set Affy Mouse430 2.0 gene chip array on mus musculus. First 10 most significant GO terms corresponding to genes cluster/group are presented in a resulting GO-analysis table.

The following figure summarizes the workflow for the analysis:

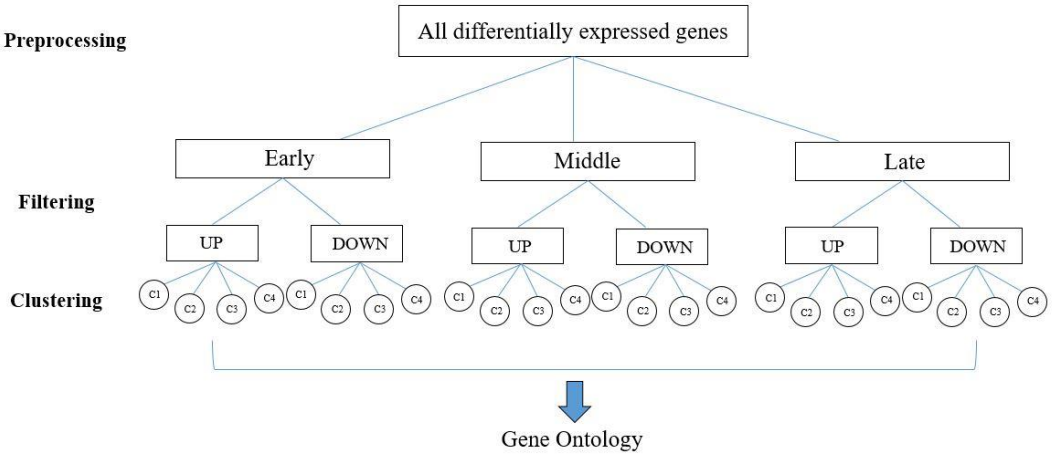

Fig1: Framework of microarray gene analysis proposed in this study

4 Results

4.1 Filtered Genes: Up-regulated and down-regulated in different intervals

Here, we present the dynamics of genes filtered in different groups based on time and expression levels. All figures for filtered genes in different groups can be found in the Appendix. For the purposes of clustering and cluster comparison, normalized intensity values are plotted across time for all the filtered genes in each interval. Figure 2 shows the dynamics of Genes up-regulated in the early inflammatory stage. It can be observed that there are indeed a range of dynamics among the genes showing up regulation: Genes up-regulated 0-6hrs and turned off soon after, Genes up-regulated 6-12hrs and turned off soon after, Genes up-regulated around 24hrs and gradually lose expression and Genes having sustained expression levels after being up regulated.

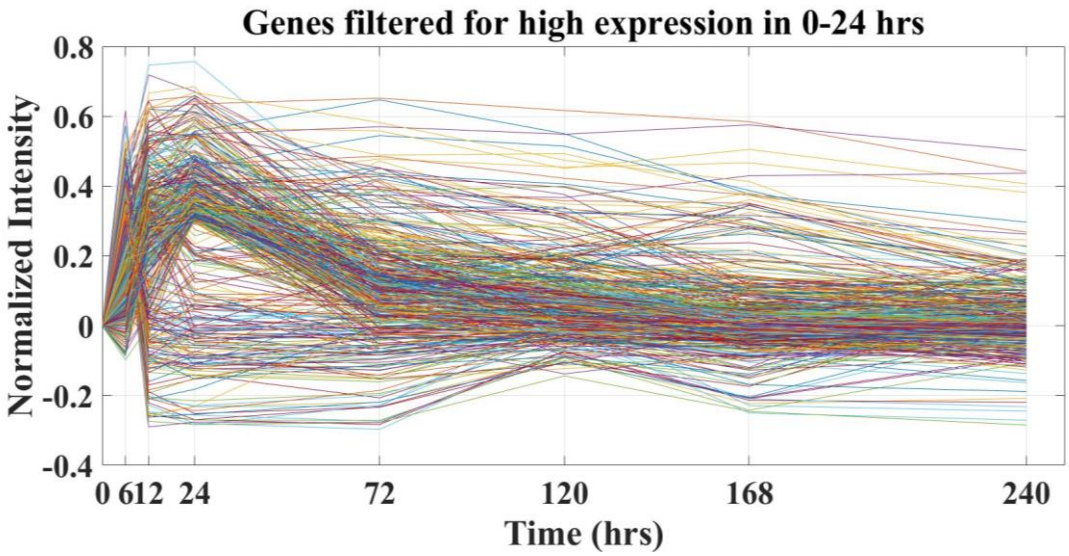

Fig2. Skin tissue : Dynamics of Genes up-regulated in 0-24hrs post wounding

4.2 Clustering Genes in each group

The filtered genes are then clustered using MATLAB's k-means clustering algorithm with k=4 to group genes having similar dynamics. The centroids of the clusters (bold lines in Fig 3) clearly depict the four groups mentioned above. The functional groups specific to these gene groups can be considered as definitive biomarkers for the early inflammatory phase of wound healing. This analysis is performed for all six filtered groups and the results can be found in the Appendix.

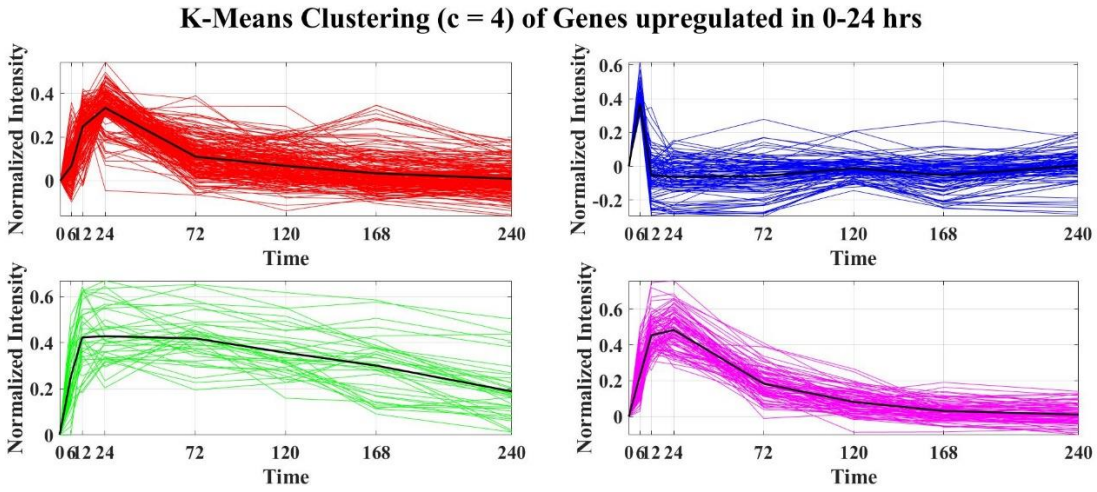

Fig 3: Skin tissue : k-means clustering of Genes up regulated in 0-24 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 219 genes, Cluster 2 (blue) – 107 genes, Cluster 3 (green) – 36 genes and Cluster 4 (magenta) – 89 genes

4.3 Differences between skin and tongue tissues

Clustering of genes completed in this work may seem non-perfect – some clusters look the same. However, with cluster representation we can see the differences between genes dynamics in both tissues. First we look at the clusters of skin and tongue tissues in 0-24 hr period. Below are the clusters obtained for tongue tissue:

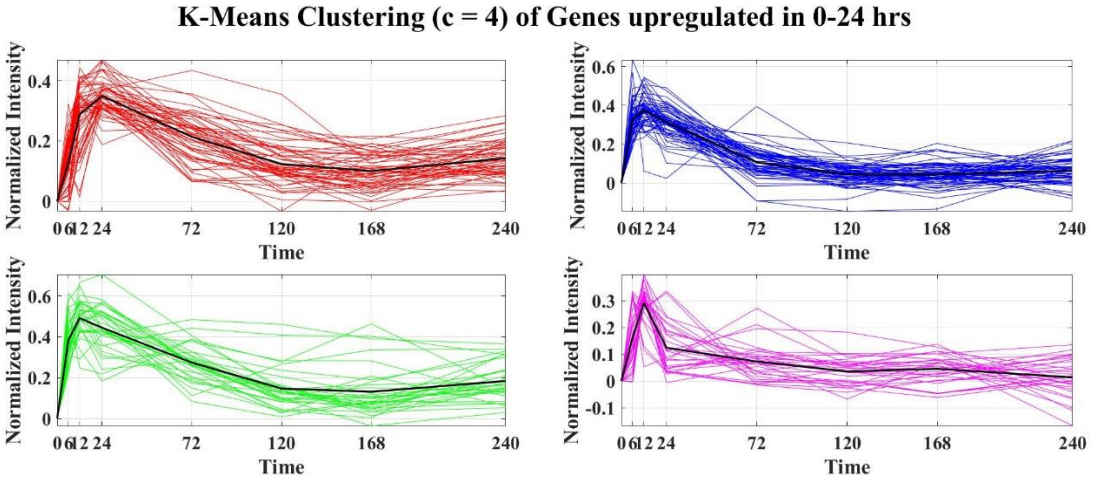

Fig 4. Tongue tissue : k-means clustering of Genes up regulated in 0-24 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 53 genes, Cluster 2 (blue) – 67 genes, Cluster 3 (green) – 32 genes and Cluster 4 (magenta) – 30 genes

- Skin Cluster 2 (blue plots in Fig 3) shows a pike of activity at 6h, no such pike observed in tongue tissue
- Skin Cluster 3 (green plots in Fig 3) – cluster with similar dynamics not found in tongue tissue
- Tongue Cluster 4 (magenta in Fig 4) a pike of activity at 12h, no such pike observed in skin tissue

4.3.1 Clusters 1, 4 in skin and 1,2,3 in tongue tissue: Inflammatory and Immune Response:

Consider clusters 1 and 4 in skin. The dynamics of gene expression in these two clusters is very similar and they could be in the same cluster. Moreover, similar dynamics is observed in clusters 1,2, 3 in tongue. Most of those genes are responsible for inflammation. Literature suggests that cytokines and chemokines help recruit and activate inflammatory cells like macrophages, mast cells, T-cells and neutrophils. In this section we compare the inflammatory and immune responses in the two tissues using the top GO terms in the 0-24 hr interval. In the skin tissue, genes in clusters 1 and 4 in the gene group up regulated in 0-24 hrs show a strong correlation to GO term GO:0006954 and GO:0006955 which signify inflammatory response and immune response respectively. The same GO terms were found in the GO analysis of tongue tissue as well, in clusters 1, 2 and 3. However, the number of genes representing inflammatory response were found to be much higher in skin tissue compared to tongue tissue. This is in correspondence with the results obtained by Lin Chen et al. [2] where the authors conclude that the inflammatory and immune responses in skin are more dominant in skin tissues compared to tongue tissue in mus musculus.

Key genes contributing to inflammatory and immune response in skin tissue are cytokines (Ccl2, Ccl3, Ccl4, Ccl7, Ccl20, Ccr1, Ccr7, and Ccrl2), chemokines (Cxcl1, Cxcl2, Cxcl3, Cxcl5 and Cxcl10) and genes from the toll-like receptor signaling pathways (Tlr1, Tlr2, Tlr4, Tlr6 and Tlr13). In contrast, although cytokines (Ccl12, Ccl3, Ccl8 and Ccr5) and chemokines (Cxcl1, Cxcl2, Cxcl3,

Cxcl5 and Cxcl10) were found in tongue tissue as well, the toll-like receptor signaling pathway genes were entirely absent. Cytokines from the interferon and interleukin family were also found to be present (IL1b, IL6, IL7r, IL1rap, IL23a, IL18rap and IFN-α r1) in the immune response of skin tissue. Figures 5 through 8 show the comparison between the expression levels for each of the above discussed gene groups.

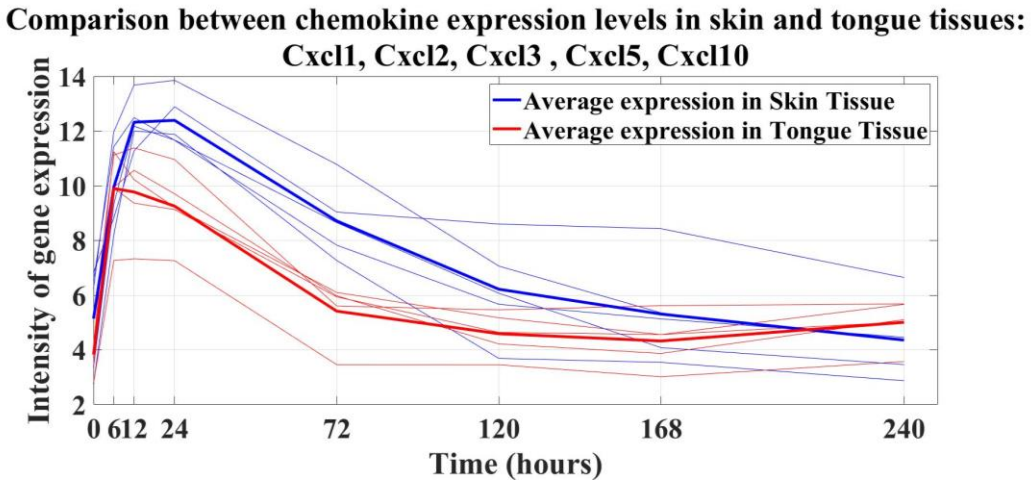

Fig 5. Chemokine expression level comparison between skin and tongue tissues

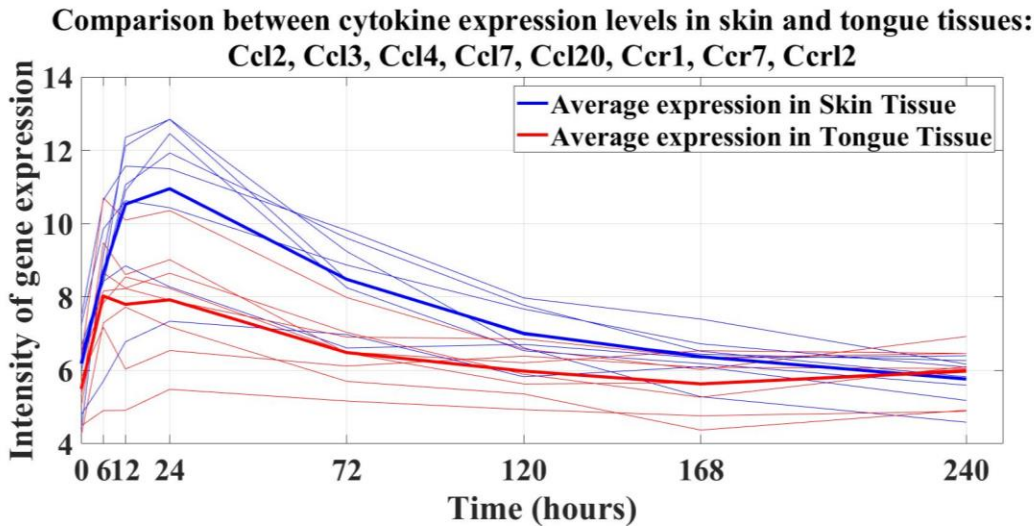

Fig 6. Cytokine expression level comparison between skin and tongue tissues

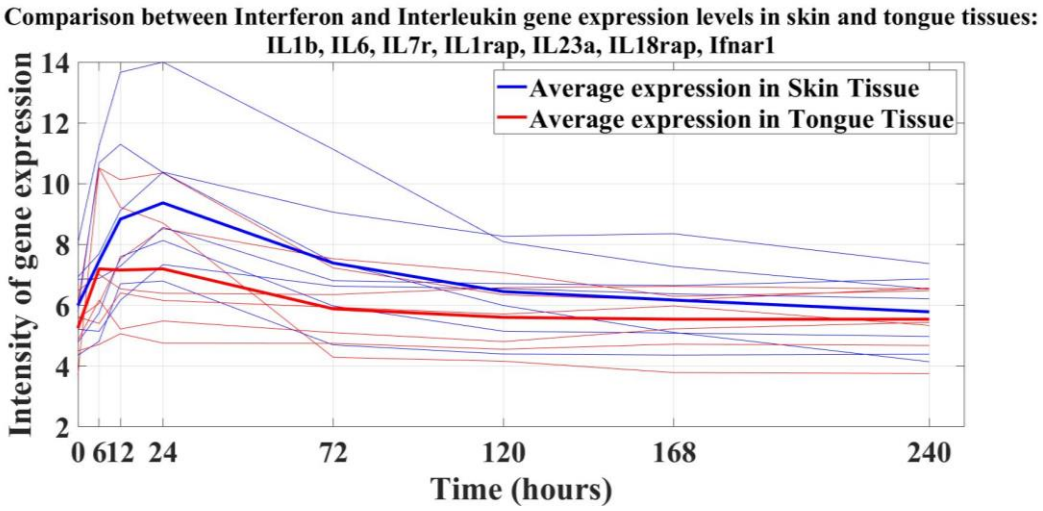

Fig 7. IL gene expression level comparison between skin and tongue tissues

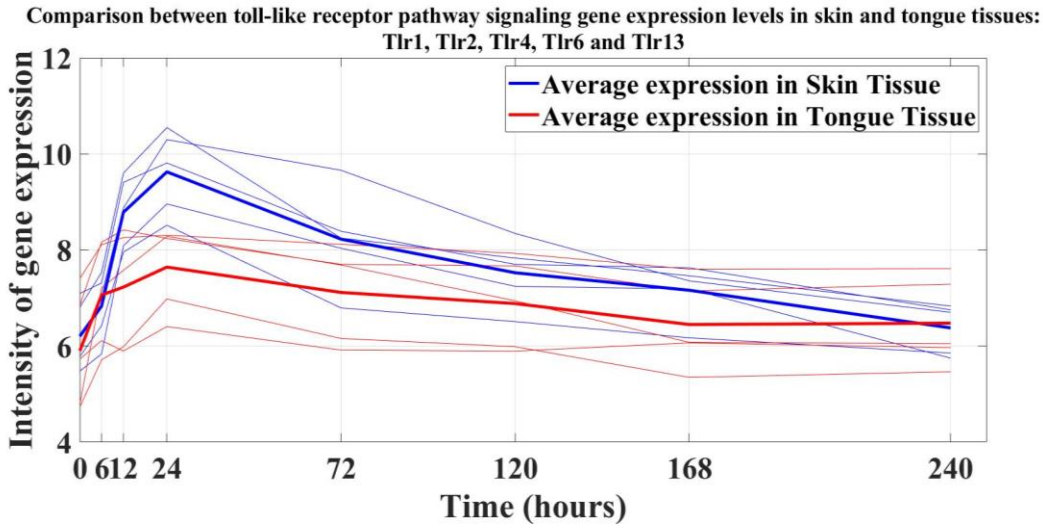

Fig 8. TLR pathway gene expression level comparison between skin and tongue tissues

One can see that in this group of genes the main difference between skin and tongue is quantitative: Gene expression in skin is higher than in tongue [2].

4.3.2 Cluster 2 in skin

The most obvious difference between skin and tongue is that skin cluster 2 has no similar dynamics in tongue. In Fig. 9 (Top) the plots of skin cluster 2 are shown both in skin and tongue. One can see that most of these genes don't show any significant dynamics in tongue but stay at low level of expression during wound healing in tongue.

Among the genes of skin cluster 2 upregulated in 0-24hrs we find an abundance of keratins and keratin associated proteins. These include Krt8, Krt18, Krt72, Krt82, Krt84, Krtap3-2, Krtap5-1, Krtap5-2, Krtap5-4, Krtap5-5, Krtap6-2, Krtap6-3, Krtap10-4, Krtap12-1, Krtap16-1, Krtap17-1, Krtap19-1, Krtap19-4, Krtap19-5, Krtap21-1, Krtap26-1, Krtap28-13 and Krtap31-1. Since keratins play a significant role in collagen synthesis, these results could explain increased scar tissue formation in skin compared to tongue tissues.

In skin cluster 3 two more genes of keratins Krt6b, Krt16 were found (Fig. 9, bottom). These genes show no differential expression in tongue tissues as well. However, Krt6b and Krt16 in tongue are initially at significantly higher level than in skin. In skin these two genes are expressed are at low level and grow after the injury. This suggests that it is not hair regeneration but their role in immune response to be the reason for high expression.

This is in correspondence with [12], paper about Krt/Krtap gene expression. The authors found that Krt6 and Krt16 were expressed differently from other keratins, concluding that high expression in non-injured tissue is observed in mouse line with better wound healing. Thus, higher expression of Krt6b and Krt16 in unwound tissue contributes to easier healing after injury.

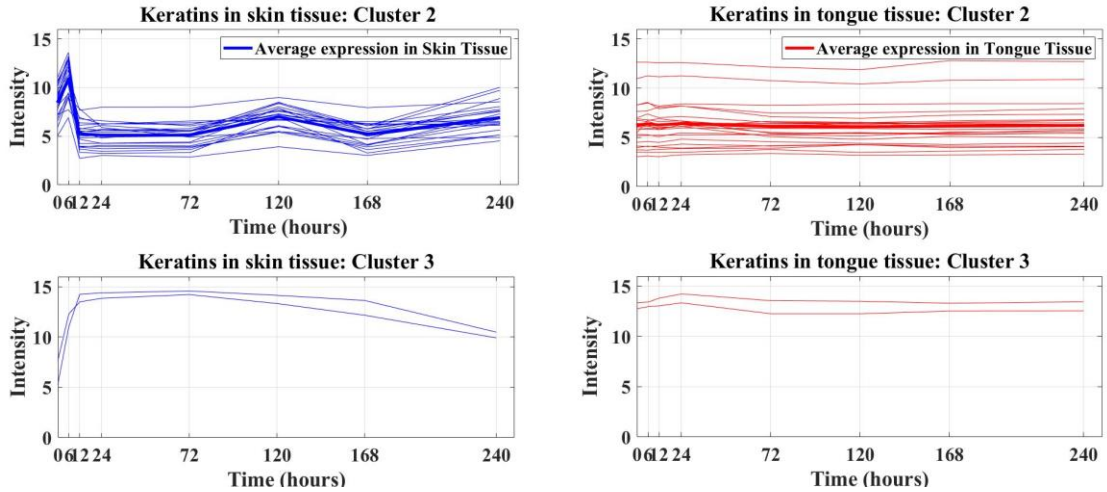

Fig 9. Genes from Skin Up regulated 0-24h Clusters 2 and 3 related to keratinocyte differentiation

Group of genes down-regulated in 0-24h interval:

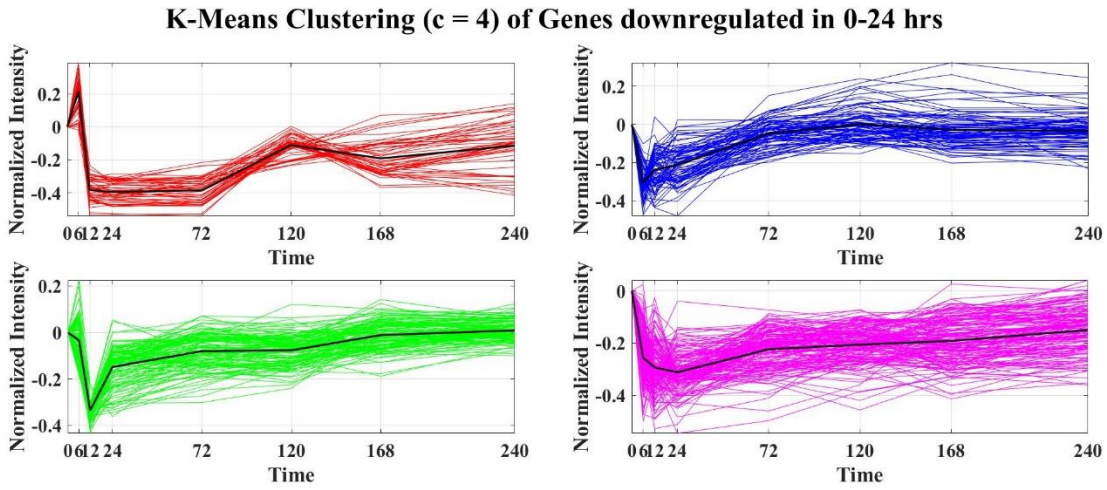

Fig 10 (a). Early down regulated genes in skin.

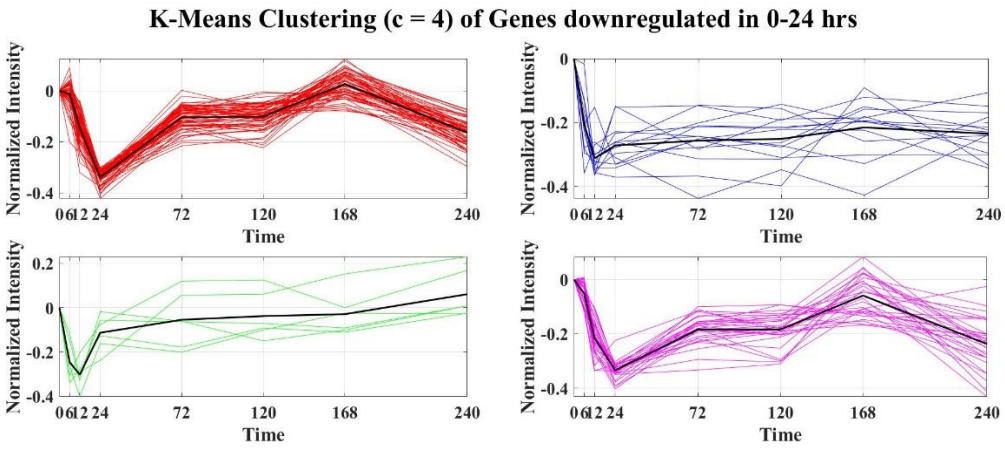

Fig 10 (b). Early down regulated genes in tongue.

- One of the differences in dynamics between skin and tongue tissue observed in Fig 10 (a) and Fig 10 (b) is a sharp peak at 6h in skin cluster 1 that doesn't exist in tongue genes.

Among the genes downregulated in the 0-24 hrs interval, we again find a few keratins in skin tissue (Krt25, Krtap6-5, Krt71, Krt27) which were entirely absent in the differential down regulated tongue tissue genes (Fig. 11).

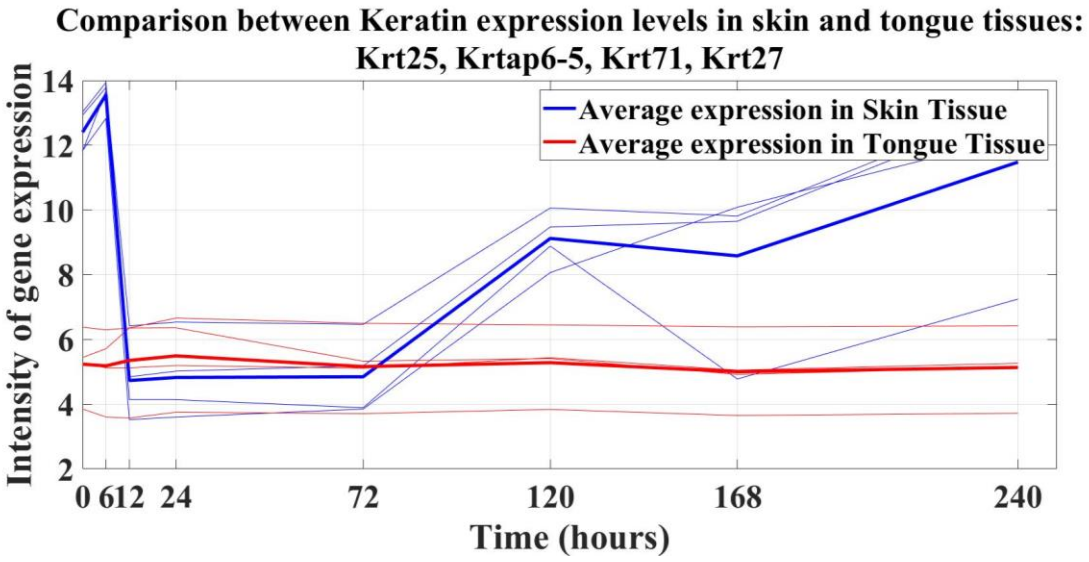

Fig 11. Down regulated Keratin gene expression level comparison between skin and tongue tissues (0-24 hrs)

It can also be seen that the skin tissue early downregulated cluster 1 has a similarity with skin upregulated cluster 2 (Fig. 9 upper).

Group of genes up-regulated in 24-120h interval:

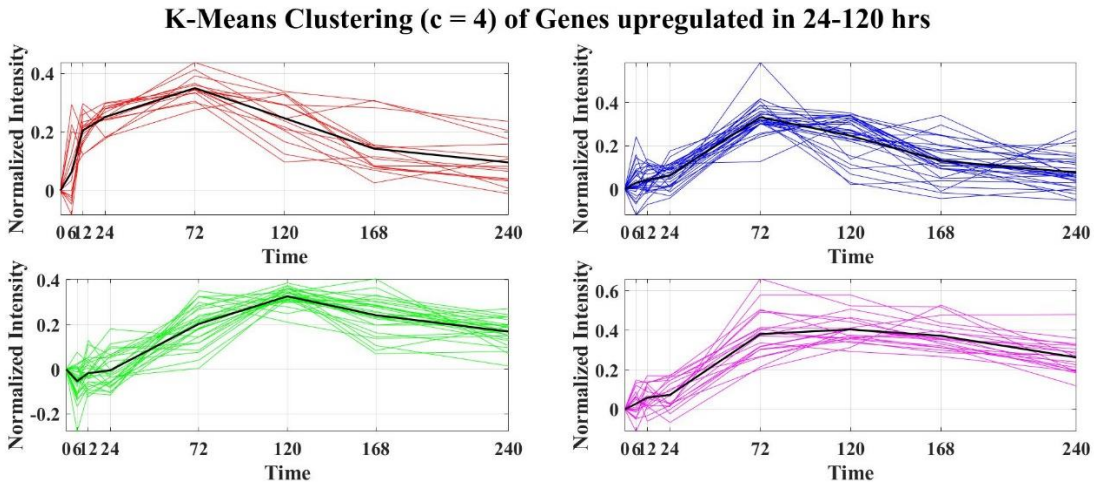

Fig 12(a). Middle-stage up regulated genes in skin

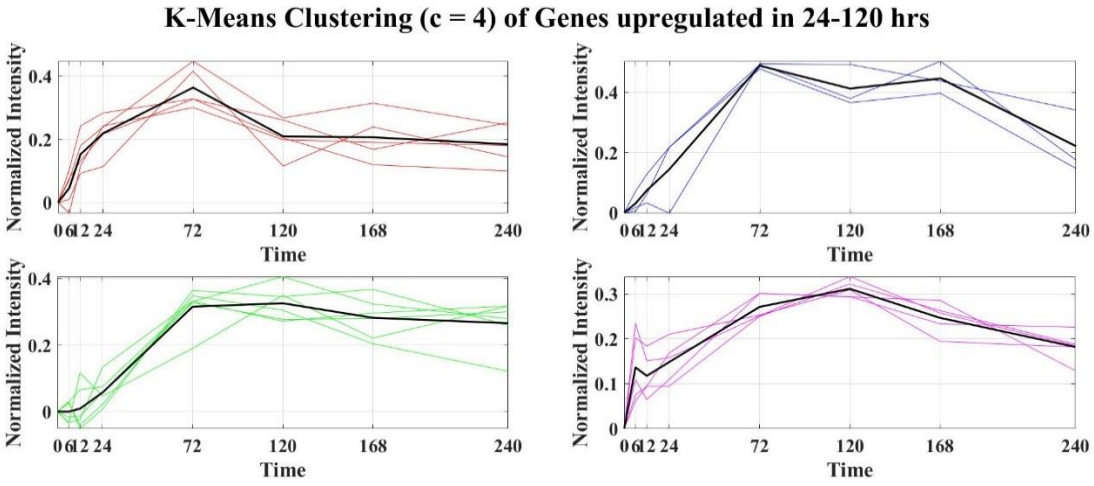

Fig 12(b). Middle-stage up regulated genes in tongue

Skin genes are more numerous, but the overall dynamics on this stage seems to be the very similar in both tissues.

Since we find fewer genes in both up and down regulated genes in the 24-120hr interval, this interval,

to obtain meaningful matches for GO terms, we combine the four clusters and look at the top 15 GO

terms in both skin and tongue. Table 5 shows a brief comparison of these terms for the two tissues. Skeletal muscle contraction, muscle tissue morphogenesis, cardiac muscle tissue morphogenesis and musculoskeletal movement standout as key processes in this interval. Troponin C type gene (Tnnc1), responsible for muscle contraction present in cardiac and skeletal muscles is found to be active in both skin and tongue tissues in this interval. In addition, Troponin I (skeletal muscle) gene which was found to be expressed in the 24-120 hr interval in skin was found to have a higher expression in the 120-240 hr interval in tongue tissue. Troponin T (cardiac muscle) gene and Actc1 (Actin alpha, cardiac muscle) a gene responsible for cardiac muscle contraction and regulation of heart function were also found to have similar dynamics in skin and tongue tissues. It should be noted that although Tnnt2 and Actc1 genes visually show similar dynamics, they weren't filtered into up regulated genes in either 24-120 hr or 120-240 hr interval for tongue tissue as their normalized intensity values were below the cut-off of 0.3 threshold which is used in this study.

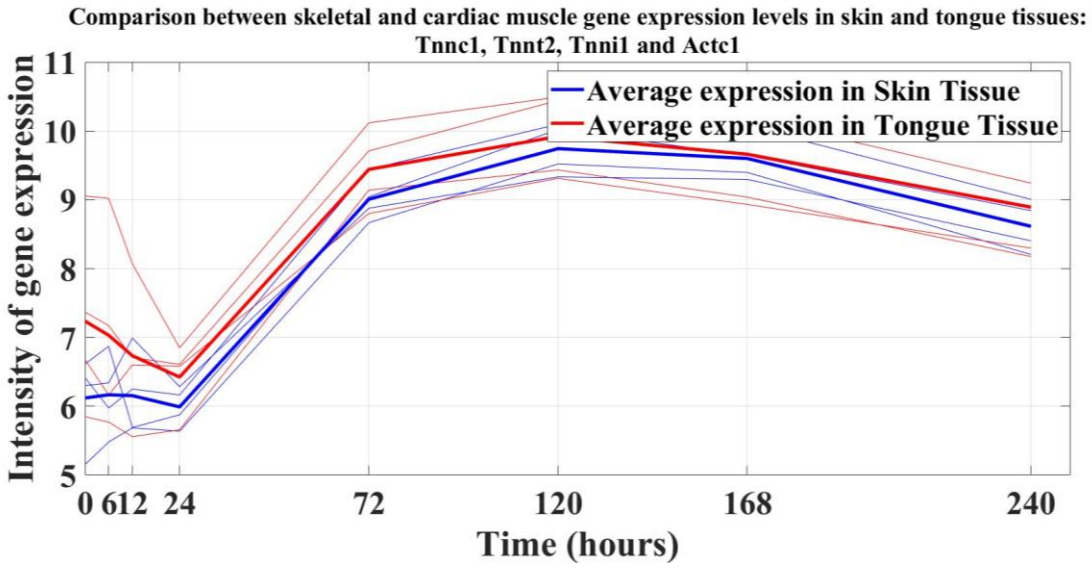

Fig 13. Skeletal and Cardiac muscle gene expression level comparison between skin and tongue tissues

Genes of this group show very similar dynamics in both tissue. This means that this stage (presumably, tissue repair) is the same in both tissue.

4.4 Gene Ontology

To identify the functionality of genes in each cluster, Gene enrichment analysis is performed. For this purpose, Web-based Gene Set Analysis Toolkit (WebGestalt) is used. The organism is chosen as *mus musculus* with the reference affymetrix chip set Affy Mouse430 2.0 gene chip array, as used in the experiment. Gene symbols for each cluster are used as the input to the software and the results for Gene Ontology with significant GO terms are obtained.

In this section, we discuss key results from the four clusters of genes up-regulated in 0-24hrs post wounding in skin tissue. A complete set of GO results for all clusters can be found in the supplementary data provided in the excel sheet

Summary_skin.xlsx and Summary_tongue_xlsx.

Tables 1-4 below show the top 10 GO terms for each cluster in Fig. 3. Consistent with literature, we find that the gene sets corresponding to defense response to wounding, immune response, inflammatory response, regulation of immune system processes etc. are found to be the most significant GO terms in the first cluster (Fig. 3: red) (Table 1). It is interesting to note that similar GO terms can be found in the fourth cluster (Fig.3 :magenta) (Table 4) as well indicating that the genes showing high regulation ~6hrs and ~12hrs could be considered in a single cluster. Gene sets significant in the second cluster (Fig. 3: blue) correspond to hormone transport, protein processing and keratinocyte differentiation which are key markers in the inflammatory phase as well.

A striking result of this type of clustering can be observed in the GO terms of the third (Fig. 3: green) cluster (Table 3) which has sustained expression levels throughout wound healing after up regulation. The most significant GO terms include skin development, tissue development, epithelium development and epidermis development which is in line with the biological understanding of the wound healing

process. It should be noted that clustering after filtering the genes into crude subgroups facilitates a deeper understanding of gene group dynamics, which would otherwise not be possible.

GO Term	Definition	p-value
GO:0006952	defense response	O(1E-16)
GO:0006955	immune response	O(1E-16)
GO:0001775	cell activation	O(1E-16)
GO:0002684	positive regulation of immune system process	O(1E-16)
GO:0002252	immune effector process	O(1E-16)
GO:0006954	inflammatory response	O(1E-16)
GO:0045087	innate immune response	O(1E-16)
GO:0002682	regulation of immune system process	4.44E-16
GO:0045321	leukocyte activation	1.55E-15
GO:0051707	response to other organism	1.88E-15

Table 1: Skin tissue : Top 10 significant GO terms for Cluster 1 of genes up regulated in 0-24hrs post wounding

GO Term	Definition	p-value
GO:0006751	glutathione catabolic process	0.000154
GO:0070327	thyroid hormone transport	0.000368
GO:0016485	protein processing	0.000657
GO:0030216	keratinocyte differentiation	0.000661
GO:0016540	protein autoprocessing	0.000922
GO:0097284	hepatocyte apoptotic process	0.001370
GO:0042219	cellular modified amino acid catabolic process	0.001716
GO:0045682	regulation of epidermis development	0.001793
GO:0051604	protein maturation	0.002057
GO:0006720	isoprenoid metabolic process	0.002242

Table 2: Skin tissue : Top 10 significant GO terms for Cluster 2 of genes up regulated in 0-24hrs post wounding

GO Term	Definition	p-value
GO:0043588	skin development	1.74961E-08
GO:0008544	epidermis development	4.68524E-08
GO:0009888	tissue development	3.52441E-07
GO:0030855	epithelial cell differentiation	1.40557E-06
GO:0060429	epithelium development	3.73959E-06
GO:0010951	negative regulation of endopeptidase activity	7.78399E-05
GO:0052548	regulation of endopeptidase activity	8.66998E-05
GO:0010466	negative regulation of peptidase activity	9.71261E-05
GO:0052547	regulation of peptidase activity	0.000124476
GO:0051047	positive regulation of secretion	0.000277611

Table 3: Skin tissue : Top 10 significant GO terms for Cluster 3 of genes up regulated in 0-24hrs post wounding

GO Term	Definition	p-value
GO:0006952	defense response	O(1E-16)
GO:0006955	immune response	O(1E-16)
GO:0006954	inflammatory response	O(1E-16)
GO:1990266	neutrophil migration	1.11022E-16
GO:0060326	cell chemotaxis	2.22045E-16
GO:0097529	myeloid leukocyte migration	2.22045E-16
GO:0030595	leukocyte chemotaxis	4.44089E-16
GO:0030593	neutrophil chemotaxis	5.55112E-16
GO:0050900	leukocyte migration	1.22125E-15
GO:0097530	granulocyte migration	1.55431E-15

Table 4: Skin tissue : Top 10 significant GO terms for Cluster 4 of genes up regulated in 0-24hrs post wounding

4.5 Comparison of GO terms in skin and tongue tissues

Skin			Tongue		
GO:0033275	actin-myosin filament sliding	4.06E-08	GO:0003009	skeletal muscle contraction	0.00031
GO:0030049	muscle filament sliding	4.54E-06	GO:0050879	multicellular organismal movement	0.00060
GO:0060415	muscle tissue morphogenesis	8.9E-06	GO:0050881	musculoskeletal movement	0.00060
GO:0003009	skeletal muscle contraction	1.04E-05	GO:0002011	morphogenesis of an epithelial sheet	0.00071
GO:0048644	muscle organ morphogenesis	1.41E-05	GO:0003208	cardiac ventricle morphogenesis	0.00118
GO:0007517	muscle organ development	3.68E-05	GO:0008015	blood circulation	0.00278
GO:0050879	multicellular organismal movement	3.75E-05	GO:0003013	circulatory system process	0.00293
GO:0050881	musculoskeletal movement	3.75E-05	GO:0002461	tolerance induction dependent upon immune response	0.00337
GO:0048729	tissue morphogenesis	4.48E-05	GO:0001844	protein insertion into mitochondrial membrane involved in apoptotic signaling pathway	0.00337
GO:0051346	negative regulation of hydrolase activity	8.52E-05	GO:0018057	peptidyl-lysine oxidation	0.00337
GO:0055008	cardiac muscle tissue morphogenesis	0.00010	GO:0003231	cardiac ventricle development	0.00352
GO:0006941	striated muscle contraction	0.00012	GO:0003206	cardiac chamber morphogenesis	0.00367
GO:0003208	cardiac ventricle morphogenesis	0.00014	GO:0072359	circulatory system development	0.00390
GO:0070252	actin-mediated cell contraction	0.00016	GO:0006941	striated muscle contraction	0.00394
GO:0044278	cell wall disruption in other organism	0.00017	GO:0002072	optic cup morphogenesis involved in camera-type eye development	0.00404

Table 5: Comparison between genes up regulated in 24-120hrs in skin and tongue tissues

5 Genes active in skin tissue and inactive in tongue tissue

There are unlimited numbers of possible differences between skin and tongue gene expression. One particular case of such difference is when one and the same gene demonstrates significant change in expression in one tissue but in another tissue the expression stays at the same level. This means that corresponding pathway in wound healing is working in one tissue but is turned off in another. In this chapter we are searching for the genes with this type of dynamics.

There are 45 genes in Skin tissue that show a normalized intensity > 0.3 at some time point between 0-240hrs which are otherwise inactive in tongue (satisfying condition (*)). These genes are filtered out and a k-means clustering is performed to identify significant GO terms related to each cluster. Fig 14 shows the results of clustering with k=9 (since there are a wide range of time series dynamics, a higher k-value is used to potentially filter all sub-groups). Clusters 2, 3 and 4 with 11, 10 and 6 genes respectively stand out as key identifiers of GO terms unique to skin tissue. Specifically, these include GO:0009913 epidermal cell differentiation (p-value 0.0018), GO:0043588 skin development (p-value 0.0032), GO:0008544 epidermis development (p-value 0.0041), GO:0042303 molting cycle (p-value 3.96E-05), GO:0042633 hair cycle (p-value 3.961E-05), GO:0031069 hair follicle morphogenesis (p-value 0.003068801) etc. which from our understanding of the biological processes are indeed specific to skin tissue. However, when a similar analysis is performed to identify Tongue genes that are not active in skin (using the same threshold of 0.3) we find that there is only one gene Vti1a.

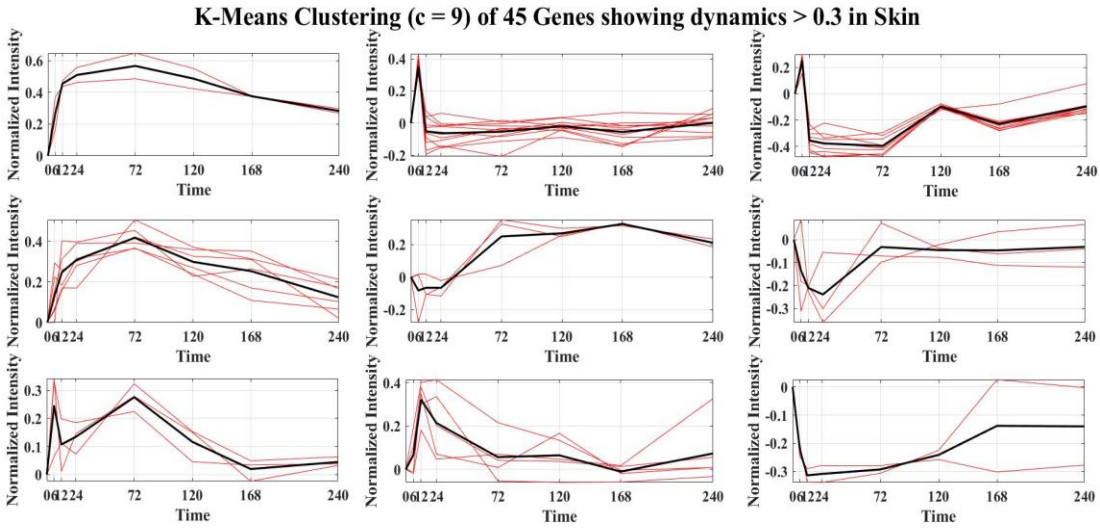

Fig 14: k-means clustering of genes active in skin which are inactive in tongue

The following table lists the top 15GO terms for the 45 genes identified above:

GO Term	Definition	p-value
GO:0042742	defense response to bacterium	3.67E-05
GO:0009617	response to bacterium	7.37E-05
GO:0051707	response to other organism	9.40E-05
GO:0043207	response to external biotic stimulus	9.63E-05
GO:0008544	epidermis development	0.000116
GO:0009607	response to biotic stimulus	0.000124
GO:0045861	negative regulation of proteolysis	0.000295
GO:0009913	epidermal cell differentiation	0.000306
GO:0010951	negative regulation of endopeptidase activity	0.000686
GO:0042303	molting cycle	0.000797
GO:0042633	hair cycle	0.000797
GO:0010466	negative regulation of peptidase activity	0.000818

GO:0043588	skin development	0.000870
GO:0050832	defense response to fungus	0.000895
GO:0098542	defense response to other organism	0.000959

Table 6: Top GO terms for the 45 genes that show activity in skin tissue which are inactive in tongue tissue

6 Discussion

Using the framework described for filtering and clustering gene microarray data, we immediately observe certain gene dynamics in skin tissue which are delayed or in some cases entirely absent in the tongue tissue. Qualitative trends of these gene are presented in section 4.3 using the centroids of the clusters obtained. Clusters 2 and 3 of the genes up regulated in the early stages of inflammation (0-24hrs) have dynamics exclusive to the skin tissue. Upon further analysis of these gene sets, we find that the sharp peak in expression at 6hrs followed by a sudden decrease in expression is due to keratins and keratin associated proteins. A majority of these keratins showed no differential expression in tongue tissue indicating that the process of keratinization and keratinocyte differentiation could be used as key biomarkers in the early inflammatory stages in skin tissue. Similarly, the genes in cluster 3 in skin tissue showed prolonged expression after up regulation, a dynamic unique to skin tissue. Among these genes, Keratin Krt16, Stefin family of genes (Stfa1, Stfa3, Stfa2l1), matrix metalloproteinases (Mmp13, Timp1) were most notable. A similar trend was also observed in keratin gene dynamics in the first cluster of the early down regulated genes where a sharp decline is observed around 6hrs. These keratins again showed no differential down regulation in tongue tissue. In addition, among the genes up regulated in the 0-24hr interval, we find numerous cytokines, chemokine, interleukins and toll-like receptor signaling pathway genes in skin tissue which presented a higher expression level compared to tongue tissues. This observation is in line with the analysis presented by Lin Chen et al. [2] and further helps validate the results in the present study. In order to identify further differences in responses to injury between the two tissues, genes showing up and down regulation in skin tissue which were otherwise inactive in tongue were analyzed. Upon investigating the absolute

gene intensity values for these genes, we find that although the genes showed no differential expression in tongue, the baseline expression intensity was higher in tongue compared to skin tissues. Fig 15 below shows a comparison between the two tissues for the set of 45 genes clustered previously.

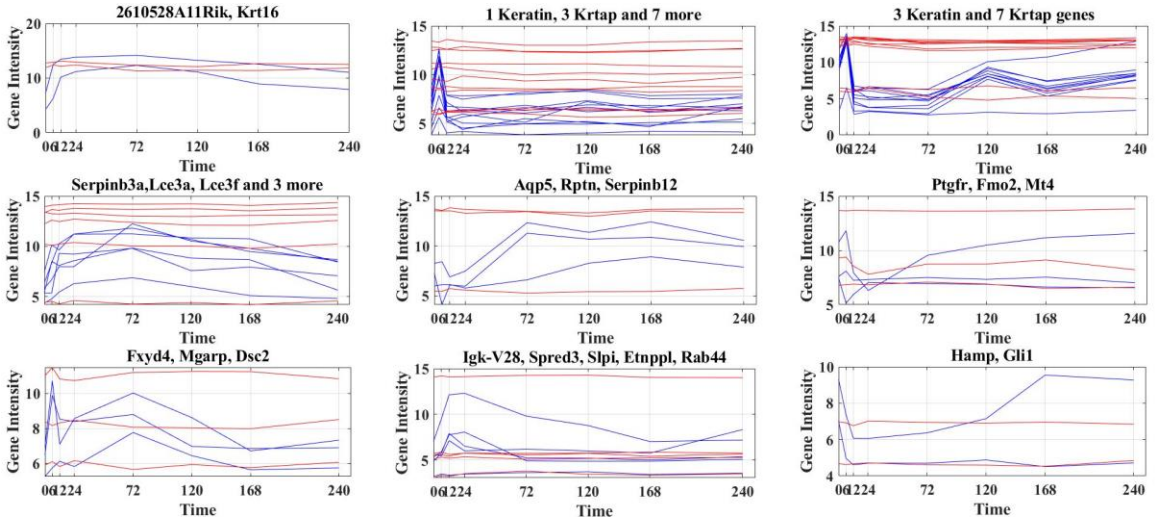

Fig 15. Absolute gene intensity comparison for the genes active in skin tissue which are inactive in tongue tissue

Among the 45 genes, 33 genes (Krtap3-2, Krtap3-3, Krtap4-1, Krtap4-6, Krtap4-7, Krtap4-13, Krtap4-16, Krtap12-1, Krtap13-1, Krt16, Krtap26-1, Krt34, Krt86, Serpinb3a, Serpinb12, Sptssb, Padi1, Defb4, Chac1, Gjb6, Wfdc12, Lce3a, Lce3f, Rptn, Fmo2, Mt4, Fxyd4, Mgarp, Dsc2, Spred3, Slpi, Rab44, 22310057N15Rik, 2310061N02Rik and 2610528A11Rik) showed a higher baseline expression in tongue tissue compared to skin tissue.

In cluster 4 of the above analysis, we find the expression of two LCE3 genes LCE3a and LCE3f. Bergboer JG et al. [22] have found that LCE3 genes, in contrast to other LCE genes, are unregulated in skin after injury and also highly expressed in skin during psoriasis. To investigate these genes further, the entire set of LCE genes are analyzed in two groups – LCE1 and LCE3 genes and their expression in the two tissues is compared. LCE3 genes indeed show the dynamics after injury as predicted in the paper [22]. In contrast to keratin 6 and 16 genes, higher LCE3 level of expression in tongue turns out to have

side-effect in the skin as their expression is related to the development of psoriasis.

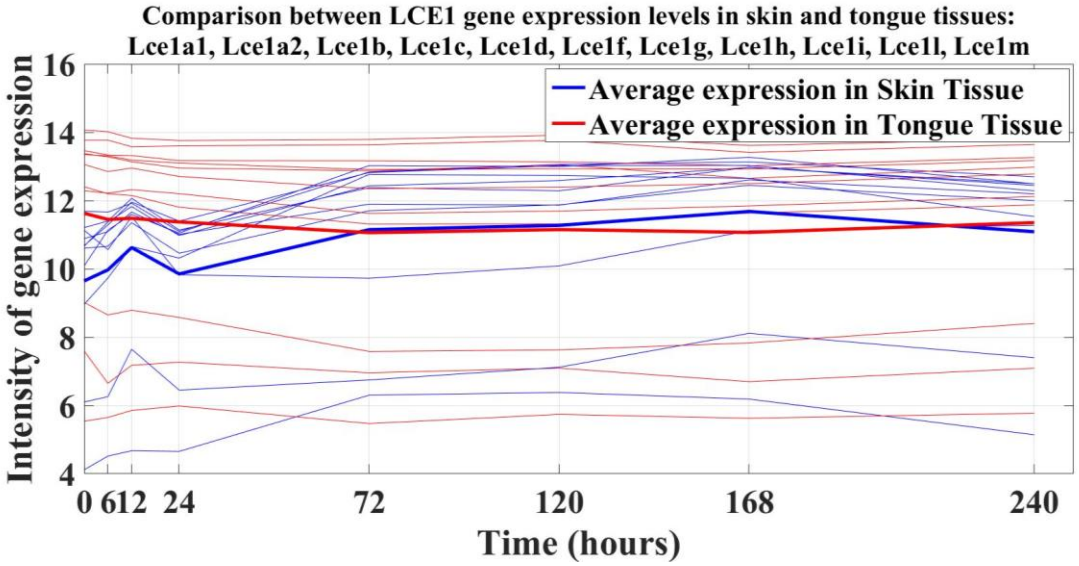

Fig 16. LCE1 gene family expression level comparison between skin and tongue tissues

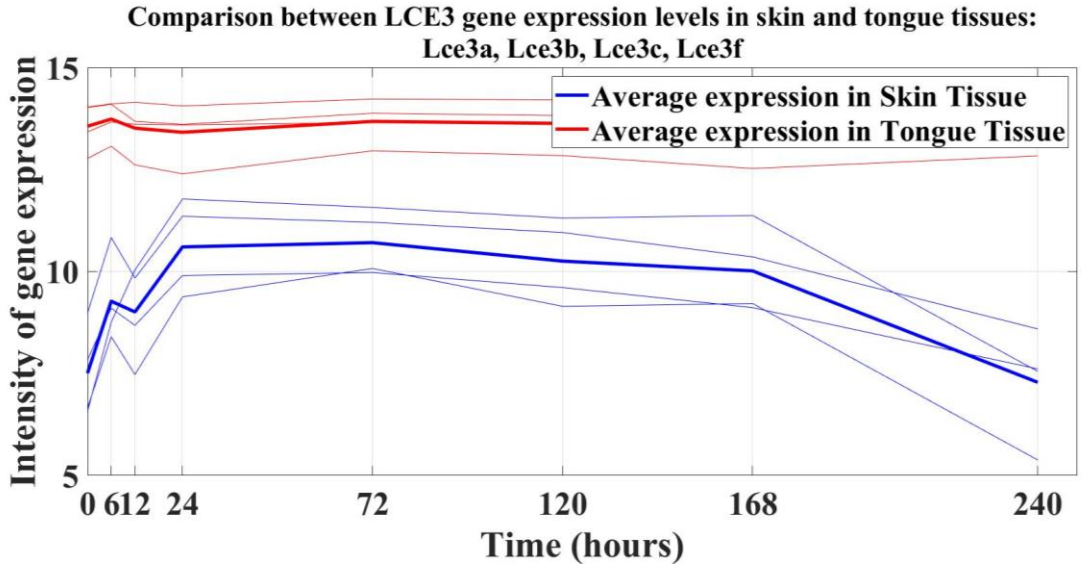

Fig 17. LCE3 gene family expression level comparison between skin and tongue tissues

Although the framework presented in this study improves the ease of clustering large data sets to obtain meaningful insights, there are a couple of drawbacks which the authors acknowledge. The choice of threshold values (0.3 for up and down regulation) is arbitrary and does influence the number of genes filtered for clustering. If there indeed exist genes which have a non-linear expression to effect dynamics, setting a cutoff for the fold intensity expression could potentially eliminate the genes from consideration. To obtain a better yield, different threshold values could be tested and a comparison could be made for the top GO terms obtained for various thresholds. Secondly, though the use of a normalized intensity value does facilitate comparison across different genes, filtering genes based on a threshold value for normalized intensity in some cases fails to filter genes which have qualitatively similar dynamics below the threshold value. An example of this issue was seen in the analysis of genes related to skeletal and cardiac muscles in section 4.3.2. Even though the genes in tongue tissue showed similar dynamics to skin tissue, they were not filtered into the final set of tongue genes as the normalized gene intensity values failed to meet the threshold of 0.3. It is our hope that these issues will be resolved in future iterations.

7 Conclusions

This study provides a framework for analyzing large gene datasets obtained for gene microarray analyses and derive meaningful observations of gene dynamics at different stage of wound healing. The results suggest an important application in identifying the specific time period of the wound healing stage based on the genetic profile of the tissue samples. Normalizing gene intensity values for comparison across genes and pre-filtering genes based on understanding of the underlying biology, facilitate effective clustering to gain detailed insights. The combination of the extracted functional groups and the top regulated genes could provide promising modeling approaches to make predictions and control the overall wound healing process of a given tissue. Having analyzed the GO sets for all clusters, we observe that this type of clustering provides a much detailed breakdown of gene dynamics compared to conventional clustering methods. Although a GO analysis of the typical five clusters approach [2] does identify similar significant gene groups, the framework presented in this work allows the identification

of specific time points where the gene regulation is significant. It is our hope that this study provides the ground work for identifying key biomarkers in developing models to predict the status of a given tissue's wound healing stage.

References

[1] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. Nature
genetics, 25(1):25, 2000.

[2] Lin Chen, Zarema H Arbieva, Shujuan Guo, Phillip T Marucha, Thomas A Mustoe, and Luisa A DiPietro. Positional differences in the wound transcriptome of skin and oral mucosa. BMC genomics, 11(1):471, 2010.

[3] Gene Ontology Consortium. The gene ontology resource: 20 years and still going strong. Nucleic acids research, 47(D1):D330–D338, 2018.

[4] James W Godwin, Alexander R Pinto, and Nadia A Rosenthal. Macrophages are required for adult salamander limb regeneration. Proceedings of the National Academy of Sciences, 110(23):9415–9420, 2013.

[5] Rajeshwar Govindarajan, Jeyapradha Duraiyan, Karunakaran Kaliyappan, and Murugesan Palanisamy. Microarray and its applications. Journal of pharmacy & bioallied sciences, 4(Suppl 2):S310, 2012.

[6] Melanie Rodrigues, Nina Kosaric, Clark A Bonham, and Geoffrey C Gurtner. Wound healing: a cellular perspective. Physiological reviews, 99(1):665–706, 2018.

[7] Piotr Andrzej Sass, Michał Dabrowski, Agata Charzynska, and Paweł Sachadyn. Transcriptomic responses to wounding: meta-analysis of gene expression microarray data. BMC genomics, 18(1):850, 2017.

[8] Chandan K Sen, Gayle M Gordillo, Sashwati Roy, Robert Kirsner, Lynn Lambert, Thomas K Hunt, Finn Gottrup, Geoffrey C Gurtner, and Michael T Longaker. Human skin wounds: a major and snowballing threat to public health and the economy. Wound repair and regeneration, 17(6):763–771, 2009.

[9] Jumaat Mohd Yussof Shah, Effat Omar, Dinker R Pai, and Suneet Sood. Cellular events and biomarkers of wound healing. Indian journal of plastic surgery: official publication of the Association of Plastic Surgeons of India, 45(2):220, 2012.

[10] John D Stroncek and W Monty Reichert. Overview of wound healing in different tissue types. Indwelling neural implants: strategies for contending with the in vivo environment, pages 3–40, 2008.

[11] Liao, Y., Wang, J., Jaehnig, E., Shi, Z., Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, Nucleic Acids Research, gkz401

[12] C.H.Cheng, J Leferovich, X.M. Zhang, K. Bedelbaeva, D. Gourevitch, C.J. Hatcher, C.T. Basson, E. Heber-Katz, K.A. Marx, Keratin gene expression profiles after digit amputation in C57BL/6 vs. regenerative MRL mice imply an early regenerative keratinocyte activated-like state// Physiol Genomics. 2013 Jun 1; 45(11): 409–421.

[13] J.Cheng, A Pipeline of Data Mining and Modeling for Mouse Wound Healing Model in Microarrary Data, Master thesis, 2019, University California in Santa Cruz

[14] Ju¨rgen Kopp et al, Accelerated Wound Healing by in Vivo Application of Keratinocytes Overexpressing KGF, DOI: https://doi.org/10.1016/j.ymthe.2004.04.016

[15] Richard J. Bodnar, Epidermal Growth Factor and Epidermal Growth Factor Receptor: The Yin and Yang in the

Treatment of Cutaneous Wounds and Cancer, DOI: 10.1089/wound.2011.0326

[16] M.P. Caley, V.L.C. Martins, E.A. O'Toole, Metalloproteinases and Wound Healing, Advances in wound care, 2015, V 4, N 4

[17] R. Iglesias-Bartolome, A. Uchiyama, A.A. Molinolo, L. Abusleme, S.R. Brooks5, J.L. Callejas-Valera, D. Edwards, C. Doci, M.-L. Asselin-Labat, M.W. Onaitis, N.M. Moutsopoulos, J. S. Gutkind, M.I. Morasso, Transcriptional signature primes human oral mucosa for rapid wound healing, Science Translational Medicine, 2018, Vol. 10, Issue 451, eaap8798

[18] P. Mostafalu, A. Tamayol, R. Rahimi, M.l Ochoa, A. Khalilpour, G. Kiaee, I.K. Yazdi, S. Bagherifard, M.R. Dokmeci, B. Ziaie, S.R. Sonkusale, A. Khademhosseini , Smart Bandage for Monitoring and Treatment of Chronic Wounds, small 2018: e1703509.

[19]Farooqui, M., Shamim, A. Low Cost Inkjet Printed Smart Bandage for Wireless Monitoring of Chronic Wounds. Sci Rep 6, 28949 (2016). https://doi.org/10.1038/srep28949

[20]J. Selberg, M. Gomez, M. Rolandi, The Potential for Convergence between Synthetic Biology and Bioelectronics, Cell Systems, 2018, https://doi.org/10.106/j.cels.2018.08.007

[21] Ojingwa J.C. and Isseroff R.R, Electrical stimulation of wound healing, DOI: 10.1046/j.1523-1747.2003.12454.x

[22] Bergboer JG et al., Psoriasis risk genes of the late cornified envelope-3 group are distinctly expressed compared with genes of other LCE groups, doi: 10.1016/j.ajpath.2010.12.017

[23] Jun-Ming Zhang and Jianxiong An, Cytokines, Inflammation and Pain, 10.1097/AIA.0b013e318034194e

[24] Kim Newton, Vishva M. Dixit, Signaling in Innate Immunity and Inflammation, 10.1101/cshperspect.a006049

Appendix

Skin Tissue Figures:

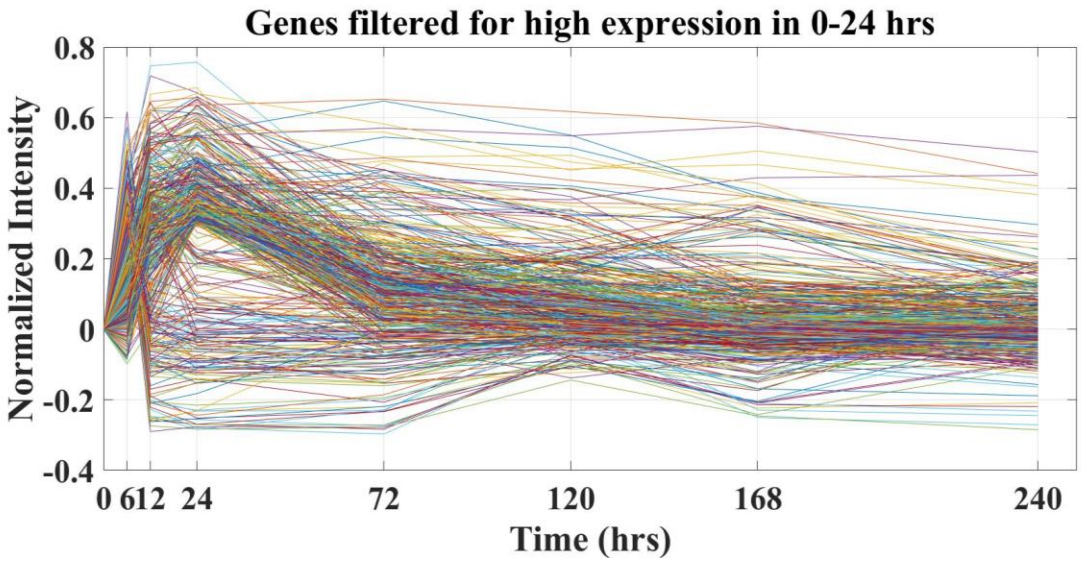

Fig A1: Skin tissue : Dynamics of Genes up-regulated in 0-24hrs post wounding

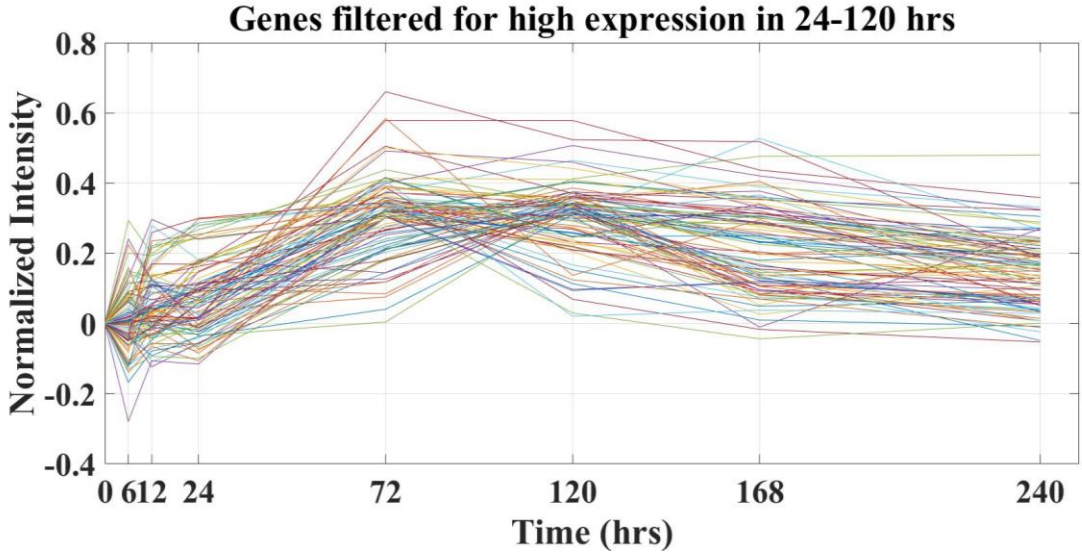

Fig A2: Skin tissue : Dynamics of Genes up-regulated in 24-120hrs post wounding

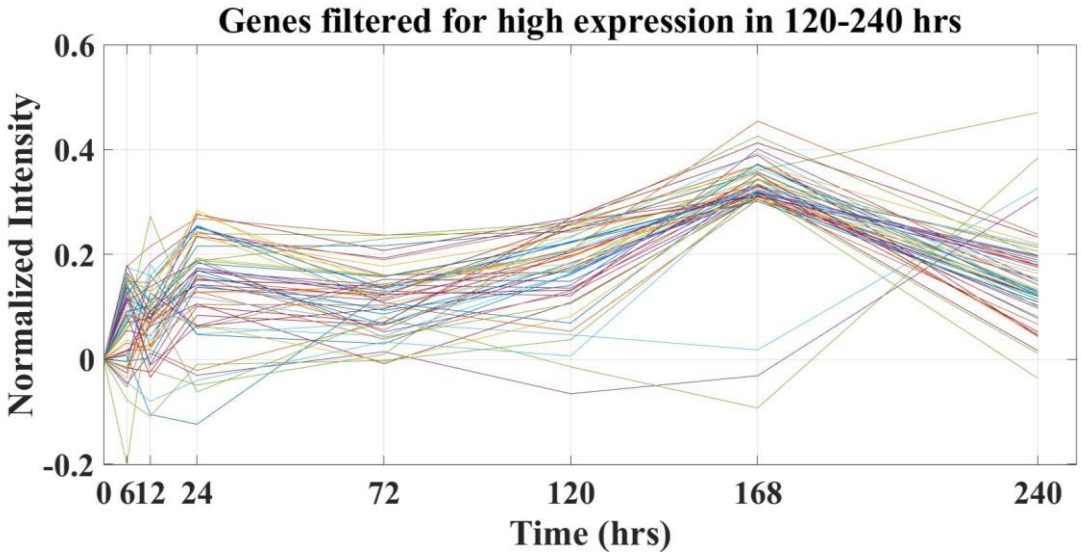

Fig A3: Skin tissue : Dynamics of Genes up-regulated in 120-240hrs post wounding

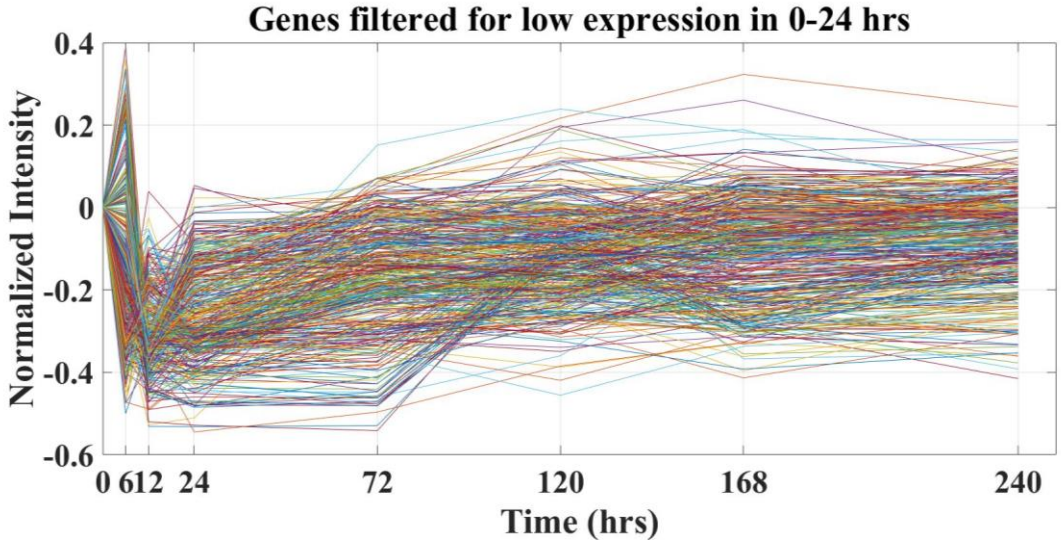

Fig A4: Skin tissue : Dynamics of Genes down-regulated in 0-24hrs post wounding

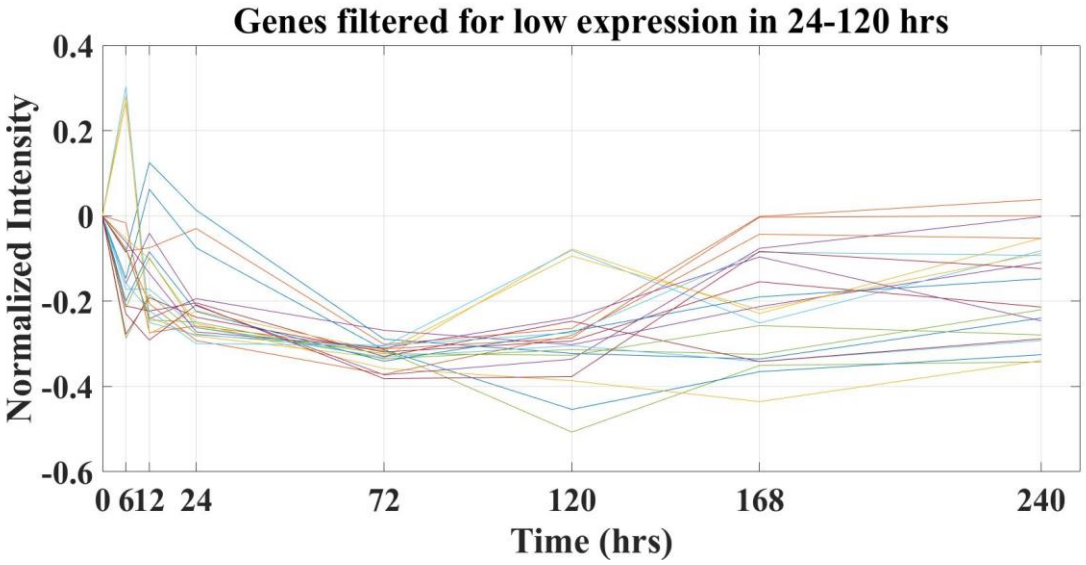

Fig A5: Skin tissue : Dynamics of Genes down-regulated in 24-120hrs post wounding

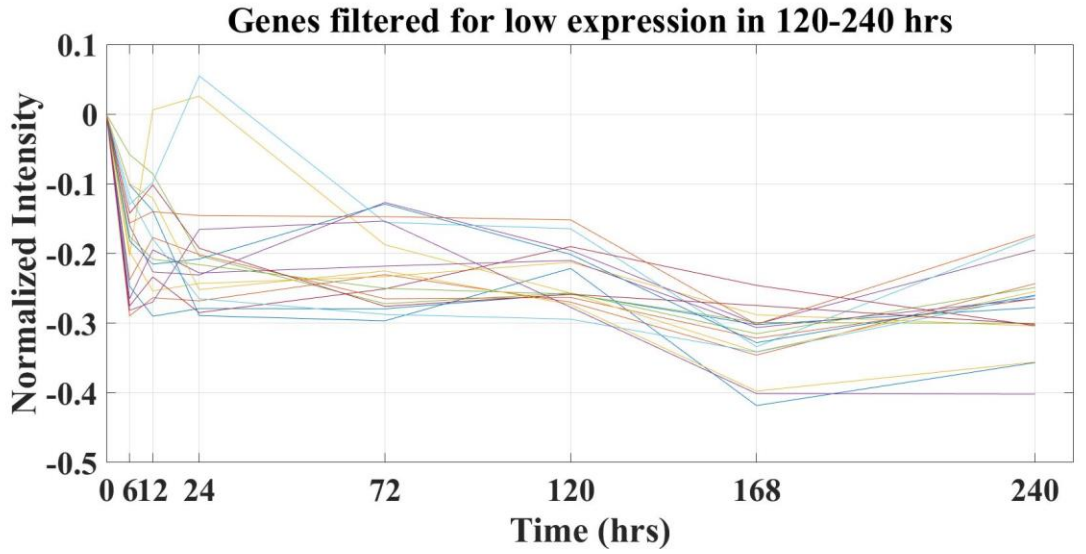

Fig A6: Skin tissue : Dynamics of Genes down-regulated in 120-240hrs post wounding

Skin Tissue Clustering Plots:

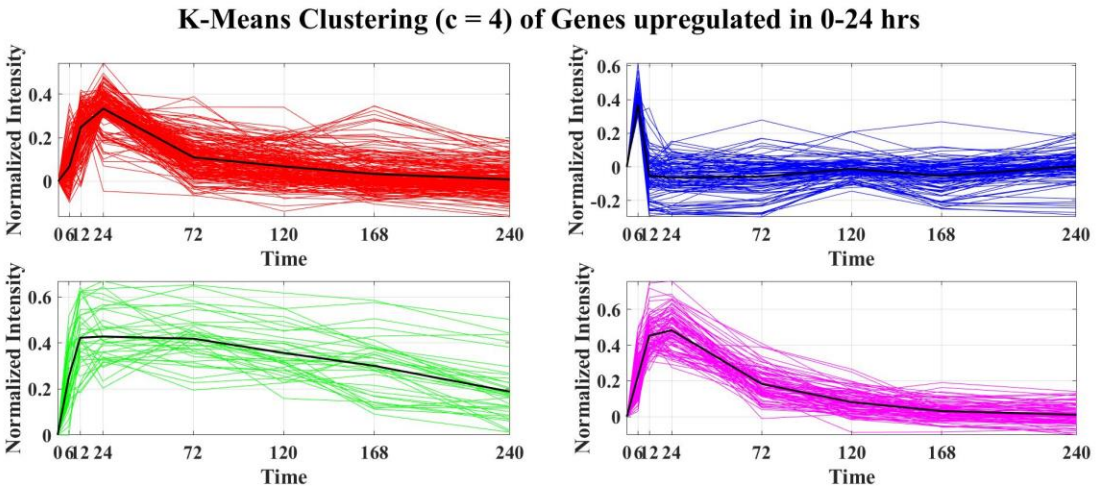

Fig A7: Skin tissue : k-means clustering of Genes up regulated in 0-24 hrs. Black lines show the centroids of each cluster.

Cluster 1 (red) – 219 genes, Cluster 2 (blue) – 107 genes, Cluster 3 (green) – 36 genes and Cluster 4 (magenta) – 89 genes

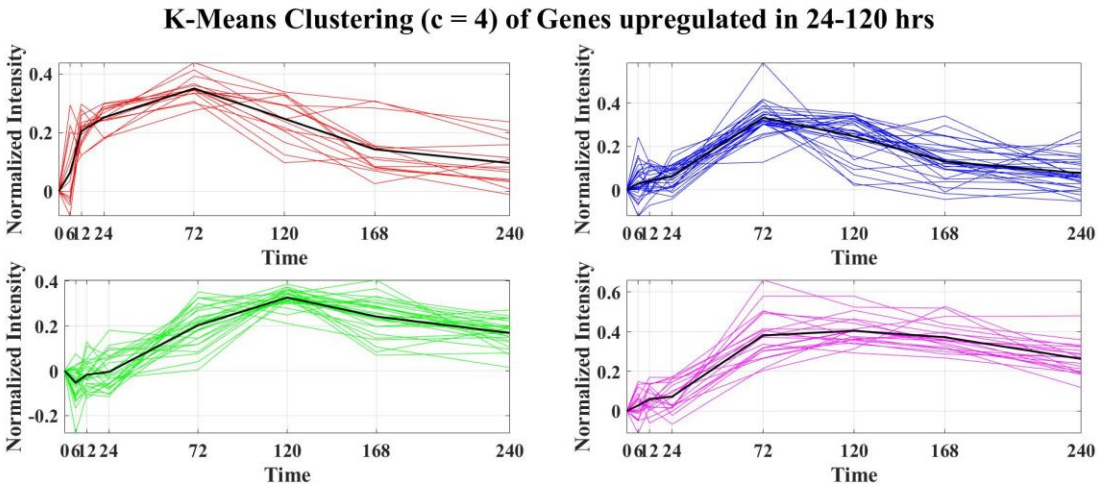

Fig A8: Skin tissue : k-means clustering of Genes up regulated in 24-120 hrs. Black lines show the centroids of each cluster.

Cluster 1 (red) – 15 genes, Cluster 2 (blue) – 31 genes, Cluster 3 (green) – 28 genes and Cluster 4 (magenta) – 20 genes

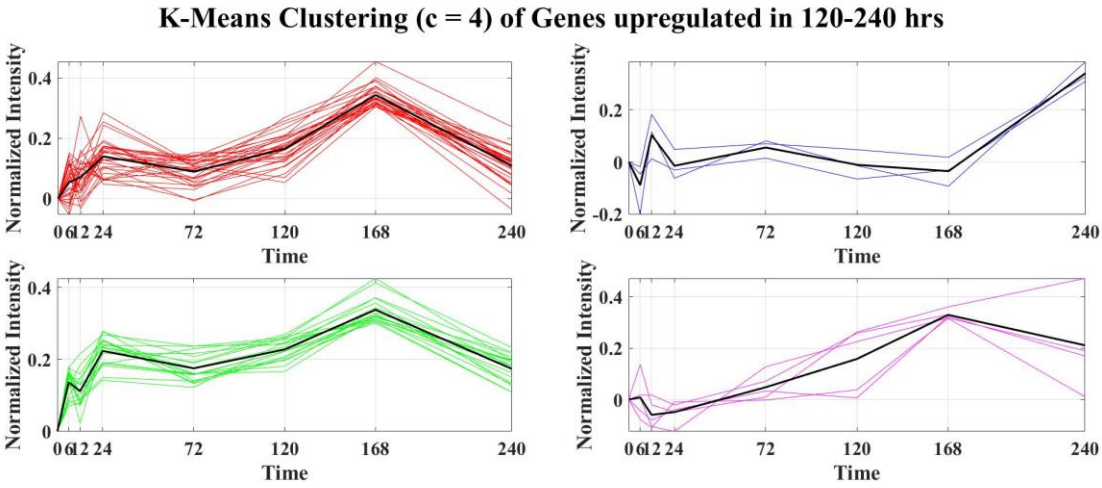

Fig A9: Skin tissue : k-means clustering of Genes up regulated in 120-240 hrs. Black lines show the centroids of each cluster.

Cluster 1 (red) – 31 genes, Cluster 2 (blue) – 3 genes, Cluster 3 (green) – 18 genes and Cluster 4 (magenta) – 5 genes

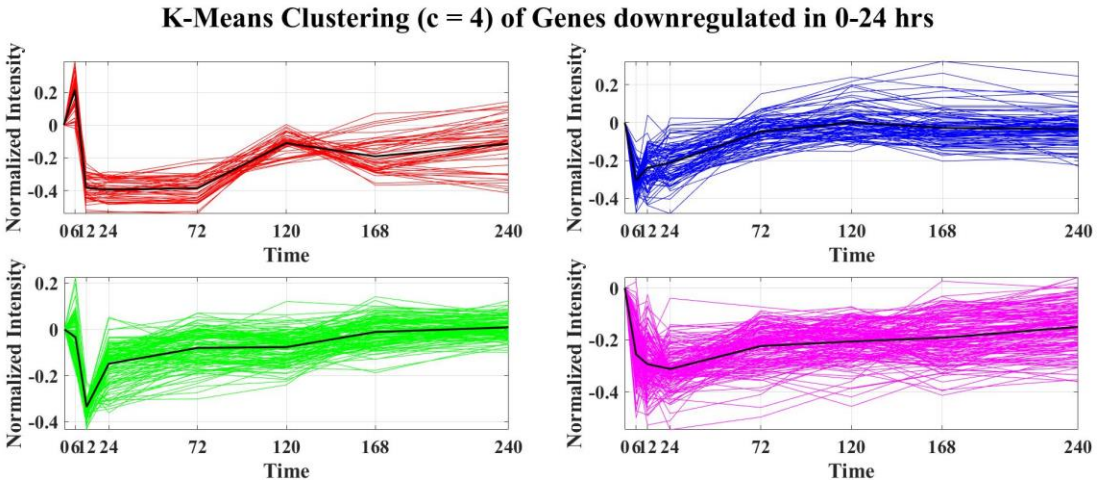

Fig A10: Skin tissue : k-means clustering of Genes down regulated in 0-24 hrs. Black lines show the centroids of each cluster.

Cluster 1 (red) – 50 genes, Cluster 2 (blue) – 91 genes, Cluster 3 (green) – 122 genes and Cluster 4 (magenta) – 163 genes

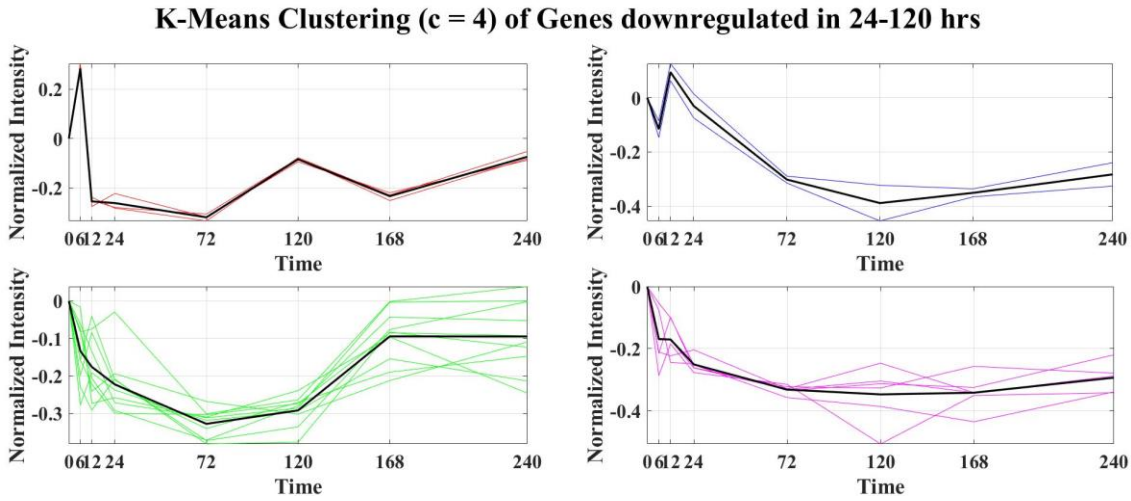

Fig A11: Skin tissue : k-means clustering of Genes down regulated in 24-120 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 3 genes, Cluster 2 (blue) – 2 genes, Cluster 3 (green) – 10 genes and Cluster 4 (magenta) – 6 genes

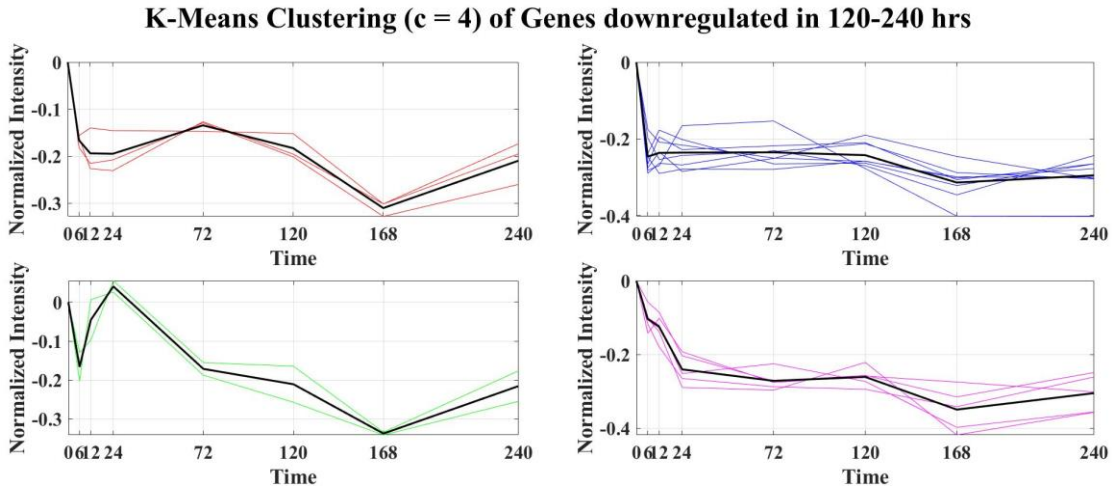

Fig A12: Skin tissue : k-means clustering of Genes down regulated in 120-240 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 3 genes, Cluster 2 (blue) – 8 genes, Cluster 3 (green) – 2 genes and Cluster 4 (magenta) – 5 genes

Tongue Tissue Figures:

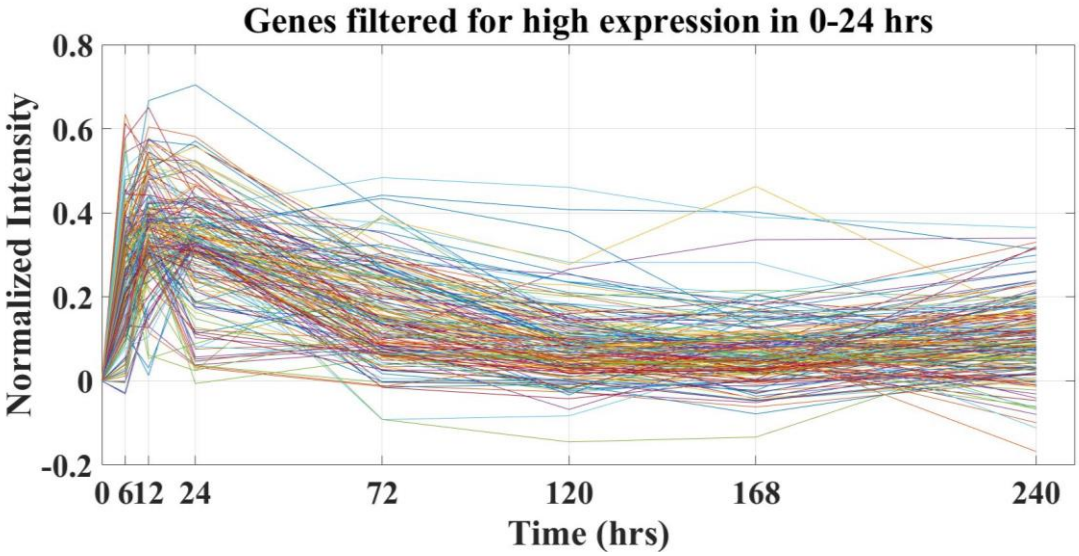

Fig A13: Tongue tissue : Dynamics of Genes up-regulated in 0-24hrs post wounding

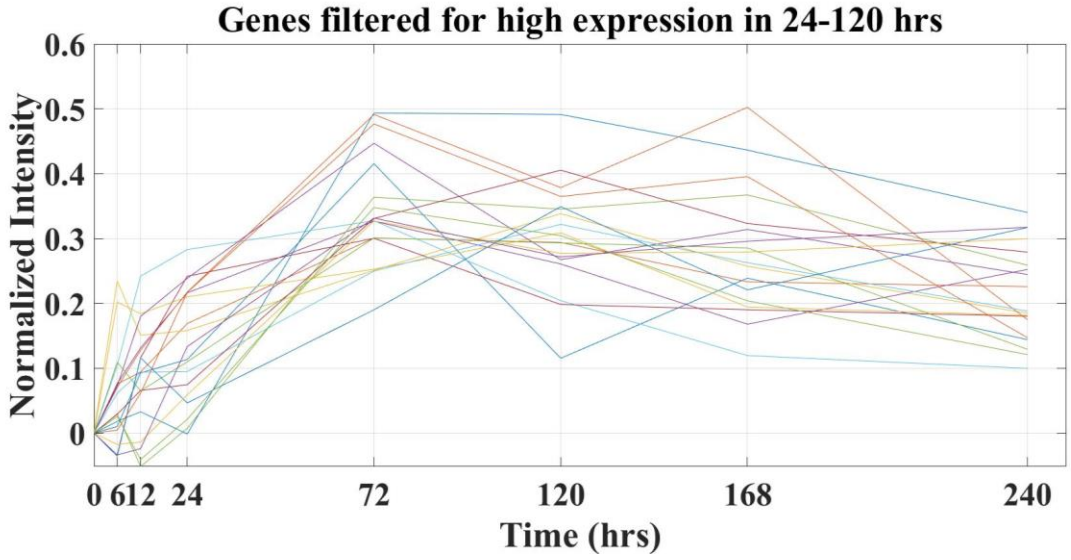

Fig A14: Tongue tissue : Dynamics of Genes up-regulated in 24-120hrs post wounding

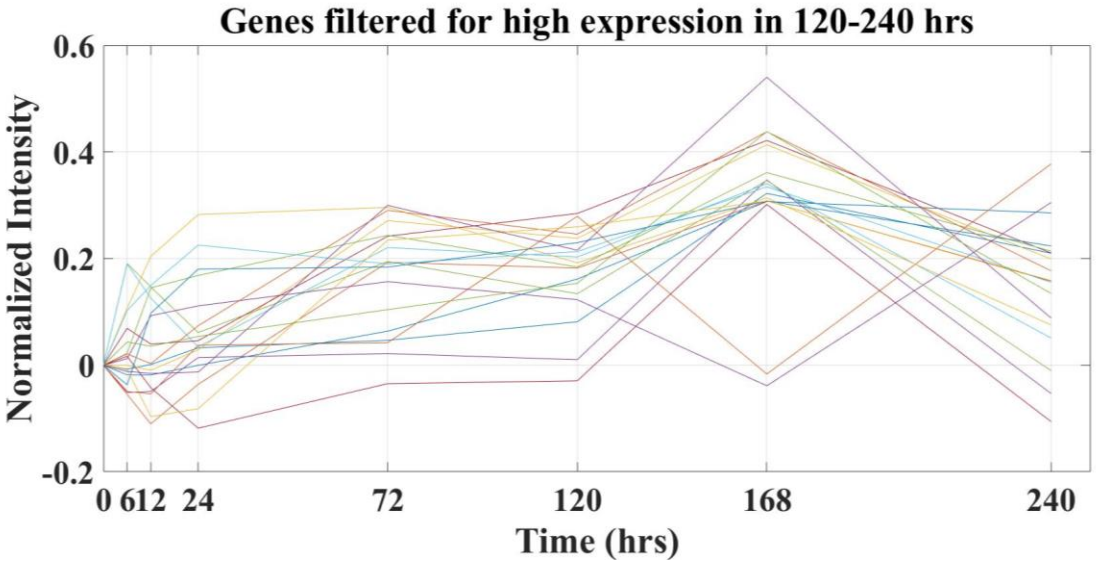

Fig A15: Tongue tissue : Dynamics of Genes up-regulated in 120-240hrs post wounding

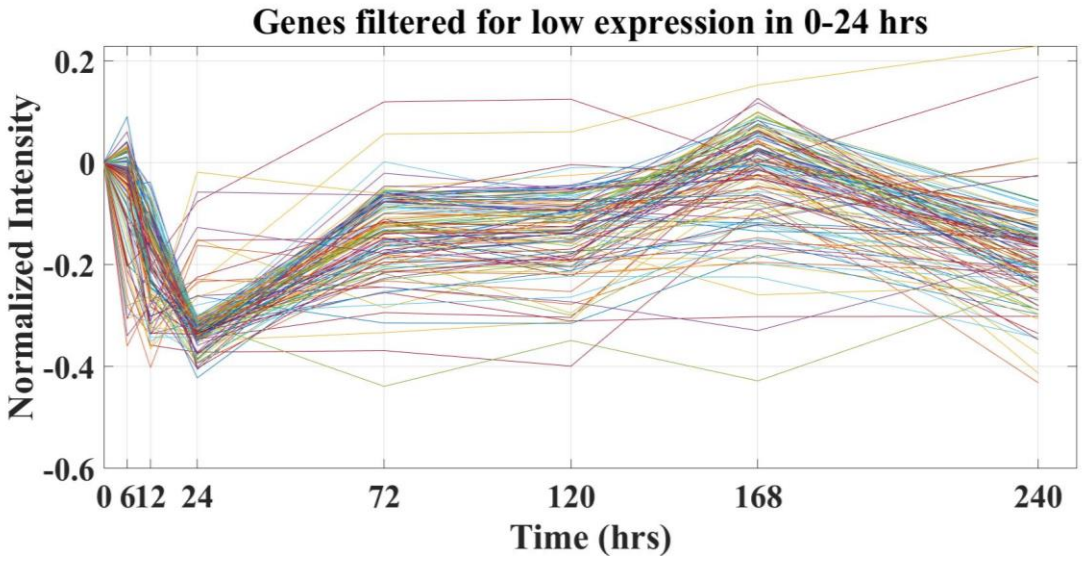

Fig A16: Tongue tissue : Dynamics of Genes down-regulated in 0-24hrs post wounding

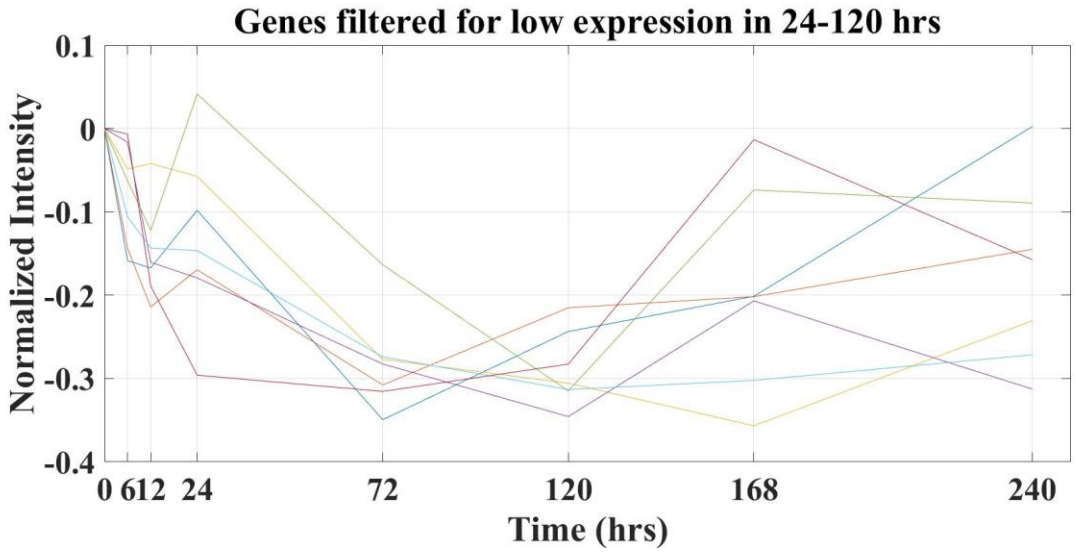

Fig A17: Tongue tissue : Dynamics of Genes down-regulated in 24-120hrs post wounding

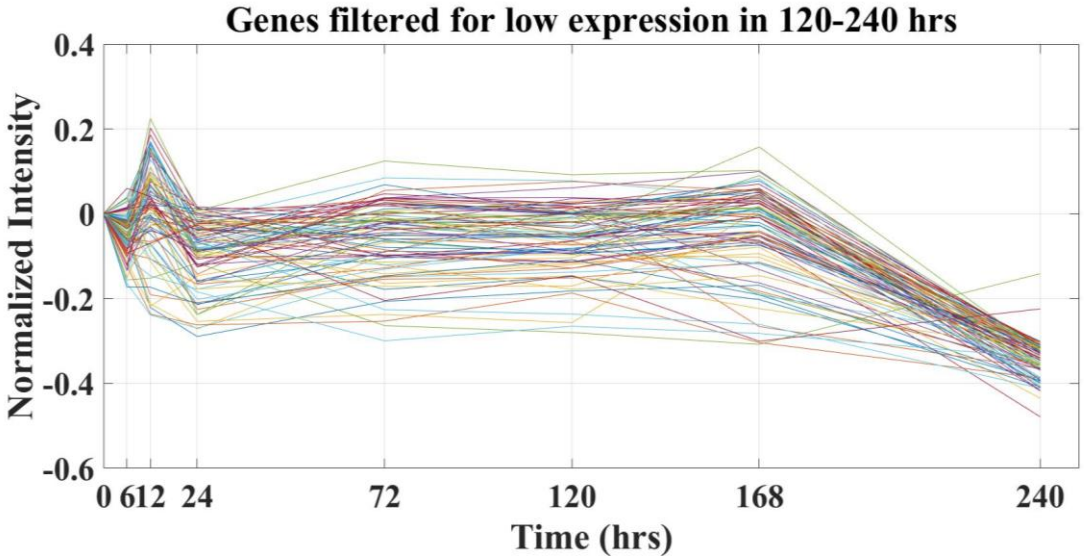

Fig A18: Tongue tissue : Dynamics of Genes up-regulated in 120-240hrs post wounding

Tongue Tissue Clustering Plots:

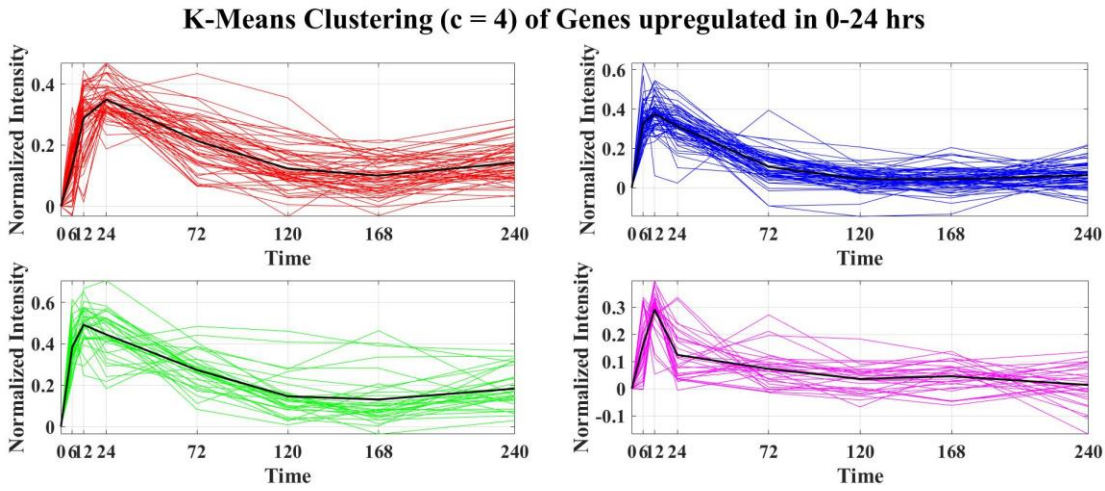

Fig A19: Tongue tissue : k-means clustering of Genes up regulated in 0-24 hrs. Black lines show the centroids of each cluster.

Cluster 1 (red) – 53 genes, Cluster 2 (blue) – 67 genes, Cluster 3 (green) – 32 genes and Cluster 4 (magenta) – 30 genes

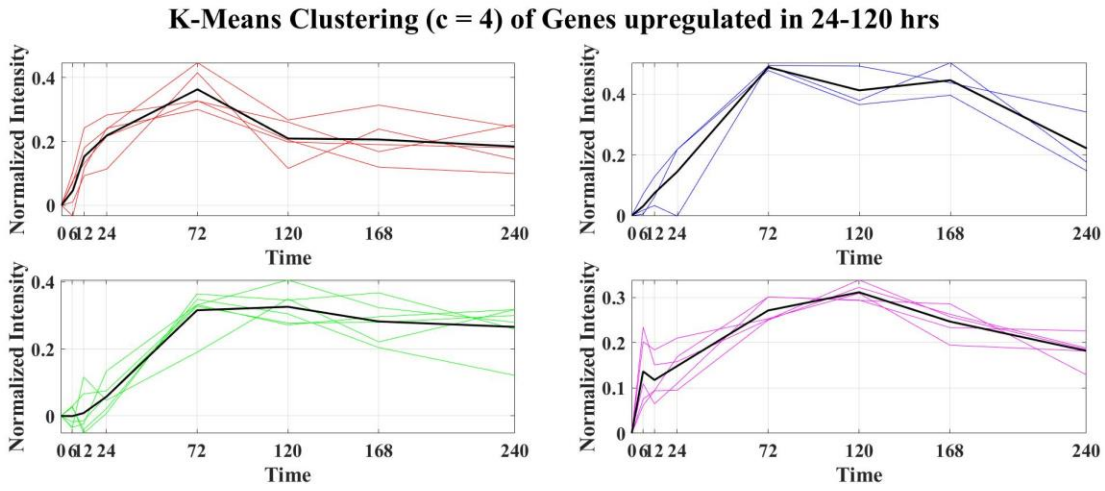

Fig A20: Tongue tissue : k-means clustering of Genes up regulated in 24-120 hrs. Black lines show the centroids of each cluster.

Cluster 1 (red) – 5 genes, Cluster 2 (blue) – 3 genes, Cluster 3 (green) –6 genes and Cluster 4 (magenta) – 5 genes

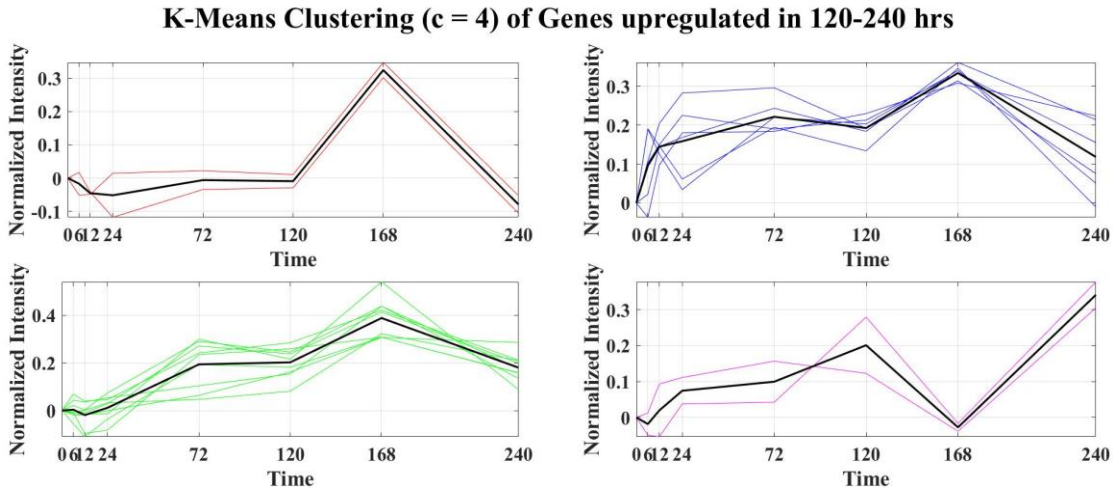

Fig A21: Tongue tissue : k-means clustering of Genes up regulated in 120-240 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 2 genes, Cluster 2 (blue) – 6 genes, Cluster 3 (green) – 9 genes and Cluster 4 (magenta) – 2 genes

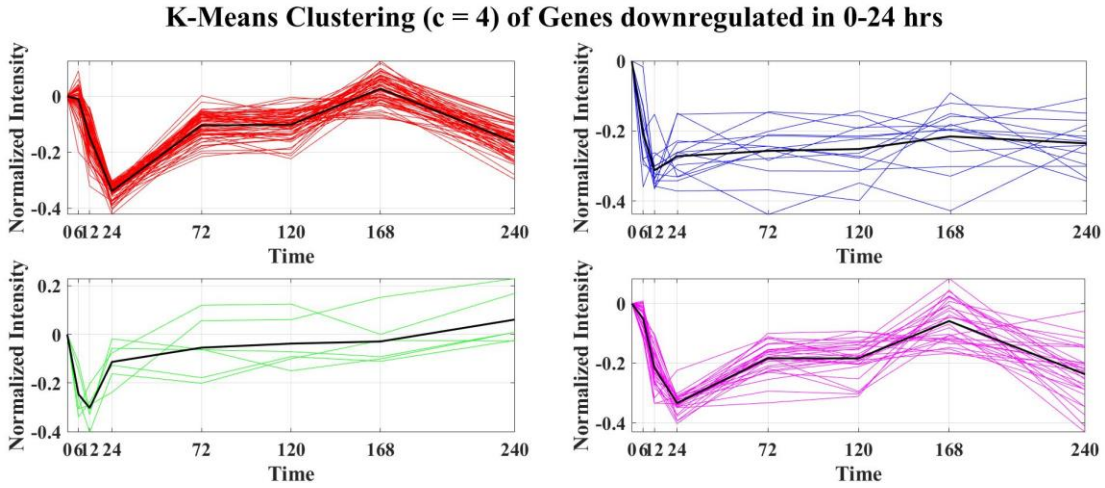

Fig A22: Tongue tissue : k-means clustering of Genes down regulated in 0-24 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 67 genes, Cluster 2 (blue) – 14 genes, Cluster 3 (green) – 6 genes and Cluster 4 (magenta) – 29 genes

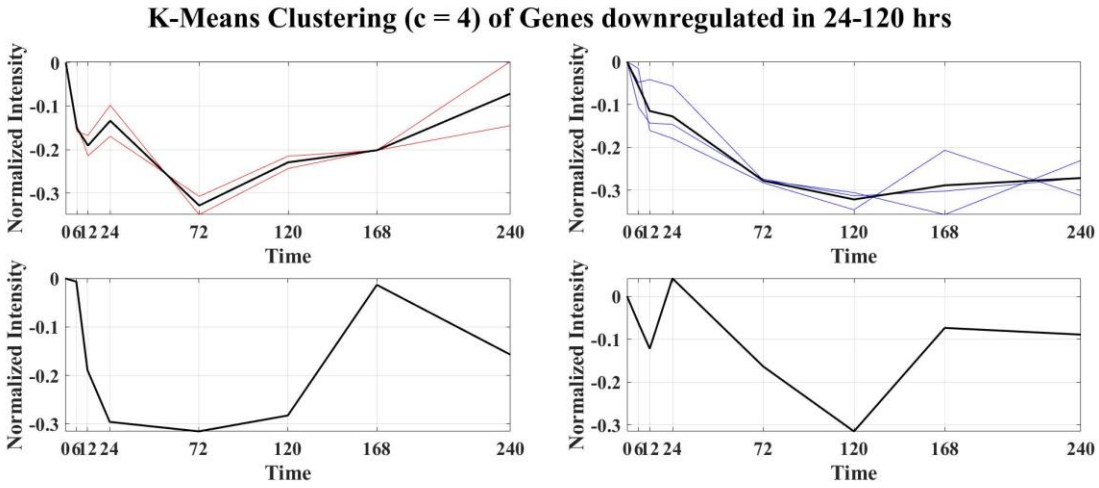

Fig A23: Tongue tissue : k-means clustering of Genes down regulated in 24-120 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 2 genes, Cluster 2 (blue) – 3 genes, Cluster 3 (green) – 1 genes and Cluster 4 (magenta) – 1 genes

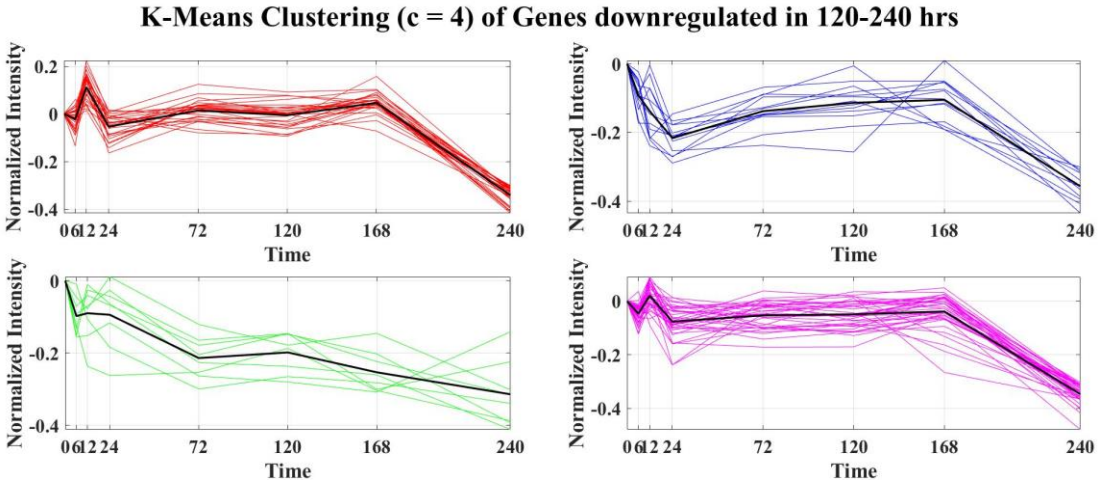

Fig A24: Tongue tissue : k-means clustering of Genes down regulated in 120-240 hrs. Black lines show the centroids of each cluster. Cluster 1 (red) – 27 genes, Cluster 2 (blue) – 13 genes, Cluster 3 (green) – 8 genes and Cluster 4 (magenta) – 37 genes