

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 October 2003 (30.10.2003)

PCT

(10) International Publication Number
WO 03/089668 A2

- (51) International Patent Classification⁷: **C12Q 1/68**
- (21) International Application Number: PCT/GB03/01710
- (22) International Filing Date: 17 April 2003 (17.04.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
0209040.5 19 April 2002 (19.04.2002) GB
- (71) Applicant (*for all designated States except US*): **SINGAPORE EYE RESEARCH INSTITUTE** [SG/SG]; 11 Third Hospital Avenue, SNEC Building, 7th Floor, 168751 Singapore (SG).
- (71) Applicant (*for MG only*): **BRASNETT, Adrian, H.** [GB/GB]; Mewburn Ellis, York House, 23 Kingsway, London, Greater London WC2B 6HP (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **TAN, Donald** [SG/SG]; 49 Oei Tiong Ham Park, 267054 Singapore (SG). **BEUERMAN, Roger** [US/US]; 70 Versailles Boulevard, New Orleans, LA 70125 (US).
- (74) Agents: **BRASNETT, Adrian, H.** et al.; Mewburn Ellis, York House, 23 Kingsway, London, Greater London WC2B 6HP (GB).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: PTERYGIUM: DIAGNOSIS, MONITORING AND TREATMENT THEREOF

(57) Abstract: Methods for monitoring and diagnosing a disease or condition associated with pterygium are described. Comparing gene expression levels between normal and pterygial tissue has provided an insight into factors that may drive the growth of pterygium. Assay methods and the products thereof for use in the treatment of pterygium are provided.



WO 03/089668 A2

PTERYGIUM: DIAGNOSIS, MONITORING AND TREATMENT THEREOFBackground to the Invention

5 Pterygium is a growth disorder of the ocular surface with potentially significant effects on vision from obscuration of the visual axis, corneal scarring, glare and induced irregular astigmatism. In very severe cases, pterygium is sight threatening. It has a high incidence in many countries around
10 the world, for example in Singapore over 9% of adults over 40 are affected, and no medical treatment is available, surgery is the only course and recurrence rates may reach 80%. Around the world pterygium affects about 30 million people making it one of the most common ocular disorders in the world. Due to
15 the surface location of the disease tissue, it may be a practical target for gene therapy. Currently most pterygium patients have to endure this inflammatory state for many months or years while the growth becomes large enough for surgical excision. Surgery is currently only indicated when
20 there is impending threat to vision, as surgery has generally poor results, and this means that the majority of cases of mild to moderate pterygium currently go untreated.

A pterygium is a triangular-shaped growth consisting of bulbar
25 conjunctival epithelium and hyperproliferative, fibrovascular subconjunctival tissue primarily occurring medially and laterally in the palpebral fissure, and encroaching onto the cornea. Pterygium may be defined as primary or recurrent, the latter being a generally more aggressive lesion which may
30 rapidly occur several weeks to months after excision of a primary pterygium.

Clinically, persons with pterygium present with a nasal or

temporal, wing-shaped conjunctival mass which straddles the limbus and grows radially to encroach on the cornea.

Pterygium is usually bilateral, but often presents with asymmetrical involvement. Gradual progression of the lesion

5 results in ocular morbidity, and visual deterioration occurs due to tear film disruption from the pterygium, corneal

distortion resulting in irregular astigmatism and eventually dense scarring at the visual axis. Patients however often

present early with symptoms of dryness, irritation and

10 hyperemia, and ocular surface inflammation. Severe cases, which recur after surgical excision may result in continued

visual disability from ocular motility restriction and

diplopia, symblepharon and cicatricial lid entropion.

15 Pterygium has two forms, primary and recurrent. Recurrent

pterygium is found after surgical removal of the primary form in a high percentage of cases, up to 80%, and growth is often

very aggressive. Recurrent pterygium causes significant

20 morbidity which often exceeds the primary lesion, as it is associated with increased ocular surface inflammation,

subconjunctival scarring and fibrous traction of the globe. In advanced pterygium, conjunctivalisation and

neovascularisation of the cornea occur, with pannus formation. Surgery in these cases is more complex and hazardous due to

25 significant scarring and tissue distortion.

It is generally accepted that the structure and function of any tissue is the result of complex patterns and timing of

30 gene expression. However, diseases are probably caused by the interaction of one or more genes. Environmental factors, such as chemicals, sunlight, viruses and other factors, can

interact with genes, redirecting expression. Genes direct programs within the cell that result in the production of

proteins associated with the disease, and in fact characterize the disease phenotype. Environmental signals may cause a gene to remain "on" too long leading to the production of proteins that are typical of that disease, or turn "off" such that
5 there is a decrease in the ability of one cell to signal another cell. Keloids of the skin are an example of genes that do not turn off when expected, over producing extra cellular proteins such as elastin and collagen. Pterygium has similarities to keloids.

10 Pterygium is an example of a disease where a group of genes are turned on by a combination of environmental and possibly genetic factors. However, little is known as to what genes are expressed in pterygium.

15 The p53 tumour suppressor family of genes are a group of transcription factors that control the activation of other genes. There is good evidence that the tumour suppressor genes may be involved in pterygium formation. [Tan DT, Tang WY, Liu
20 YP, Goh HS, Smith DR, Br. J. Ophthalmol. 2000, 84:212-216].

The aetiology and pathogenesis of pterygium remains unclear. Long thought to be a degenerative process from light microscopic descriptions of elastotic degeneration, the
25 aggressive nature of disease recurrence and recent cellular and growth factor studies suggest that this is an active growth disorder. Epidemiological studies suggest an environmental basis in the form of chronic ultraviolet light exposure, with pterygium more prevalent in the tropics and in
30 areas with higher levels of ambient sunlight. Focal limbal stem cell deficiency has also been suggested, and stem cells may be damaged by UV radiation which affects DNA repair. Microsatellite instability, abnormal p53 expression and

aberrant epithelial apoptosis all support the theory of chronic UV radiation as a causative factor in pterygium. Studies of pterygium tissue demonstrate the role of TGF- β in the fibrovascular growth of pterygium, while other recent
5 observations include the expression of metalloproteinases in pterygium head fibroblasts, which may explain the stromal invasiveness of this condition. The expression of genes for proliferation, production of extra cellular matrix, inflammation have not been examined together to form a
10 detailed understanding of pterygium.

Disclosure of the Invention

At this time it is not known which genes are activated in
15 pterygium by the factors such as the p53 family. Uncovering genes that have a significant change in regulation should be important to understanding pterygium development. It is likely that the recurrent form of pterygium may have a different basis in gene expression and it would be advantageous to
20 determine the pattern of gene expression in recurrent pterygium, as the basis for changing therapeutic strategy. The onset of pterygium may be caused or enhanced by environmental factors, but the sequence of events leading to this growth disorder of the ocular surface are not known.

25 By assessing all known transcripts in a pterygial growth, we aimed to identify those factors involved in this disease. The development of new techniques such as gene microarrays allows comprehensive analysis of thousands of transcripts in the
30 tissue under investigation. Gene microarrays are now the preferred method for large-scale expression analysis. They are capable of simultaneous measurement of levels of tens of thousands of gene transcripts in a single hybridisation assay.

Incyte Pharmaceuticals Inc., have designed high intensity, two-dimensional arrays containing 9,128 annotated genes or EST clusters that are able to perform parallel measurements of thousands of human gene transcripts and ESTs in a single RNA sample.

We have used an Incyte array to investigate the expression of genes in pterygium. By comparing samples taken from a primary pterygial growth and from control conjunctiva, we have been able to use rigorous analytical tools to establish reliable data on genes which are expressed in primary pterygium.

Thus, investigation into differences in gene expression levels between normal conjunctiva and pterygial tissue has allowed us an insight into the factors that may drive the growth of the pterygium. There is a continuing need in the art to identify new targets for therapeutic intervention in conditions which are associated with abnormalities in the growth, regulation or function of conjunctival tissue. Additionally, there is a need to identify therapeutic agents with activity against such targets. Further, the use of such agents against these targets may have a value in the treatment and diagnosis of disease.

The knowledge of a number of transcripts, both of genes known as such and from ESTs, provides novel assay targets and allows the development of new therapies for disease. For example, in many clinical situations the expression of the genes we have identified may be used as markers for diagnosis or prognosis.

The present invention is primarily focussed on primary pterygium. However, it may be that the methods described herein are equally as applicable to recurrent pterygium and

reference herein to pterygium is meant to include both the primary and recurrent forms of the disease unless stated otherwise.

5 In a first aspect, the present invention provides a method of monitoring the progression of, or diagnosing, a disease or condition associated with pterygium, said method comprising:

making a quantitative determination of the transcript level of at least one gene shown in Table 1, 2 or 3 in a
10 sample of cells obtained from the site of said disease or condition; and

comparing the transcript level so determined with the transcript level of said at least one gene obtained from a control sample of cells.

15

In another aspect the invention provides a gene chip array suitable for use in the above-described method of the invention comprising at least one nucleic acid suitable for detection of at least one gene shown in Table 1, 2 or 3,
20 optionally a control specific for said at least one gene and optionally at least one control for said gene chip.

In a further aspect, the invention provides assay methods for modulators of pterygium, wherein said method comprises:

- 25 (a) providing a protein encoded by a gene selected from Table 1, 2 or 3;
- (b) bringing said protein into contact with a candidate modulator of its activity; and
- (c) determining whether said candidate modulator is capable
30 of modulating the activity of said protein;

or wherein said method comprises:

- (a) providing a pterygial cell in culture;

- (b) bringing said cell into contact with a candidate modulator of pterygium; and
- (c) determining whether said candidate modulator is capable of modulating the transcript level of at least one gene selected from Table 1, 2 or 3.

Another assay method made available by the present invention involves assays for regulators of DNA function, for example regulators of promoters or other control regions which are responsible for the production of the transcripts of Table 1, 2 or 3 herein. Thus such an assay method may comprise providing a gene transcript sequence listed herein attached to its native promoter region, bringing this into contact with a candidate modulator under conditions where in the absence of modulator gene the transcript is produced, and determining whether the candidate modulator alters transcription.

Modulators obtained by such methods may be used in a method of modulating pterygial tissue growth in a human patient.

In another aspect, the identification of ESTs has allowed new potential targets for therapeutic intervention to be developed. Thus the invention provides a vector comprising an EST sequence from Table 1 linked to a promoter for transcription of said sequence. Such vectors are useful for expression of proteins encoded by the ESTs in the analysis of the genes in pterygial growth and may have direct therapeutic use in themselves, e.g. as recombinant proteins or in gene therapy applications.

Tables

Table 1 shows the genes which are up-regulated in pterygium,

these genes being identified after comparison on the chip of pooled RNA from diseased and control tissue. Gene determination was made using the Incyte chip. In Table 1, "Diff Expr" stands for differential expression and this column shows the increase in gene expression in pterygium over control tissue. The remaining columns show the Gene ID allocated by Incyte, the Incyte Clone ID, GenBank ID and GenBank Accession Number. The gene name and locus are also provided.

Table 2 shows the genes which are down regulated in pterygium when compared to control conjunctival tissue taken from the same patient. The probe set used to obtain these results was the Affymetrix U133A. In Table 2, GO stands for Gene Ontology and GenMapp is the netaffx database from Affymetrix.

Table 3 shows the genes which are up regulated in pterygium when compared to control conjunctival tissue taken from the same patient. The probe set used to obtain these results was the Affymetrix U133A. In Table 3, GO stands for Gene Ontology and GenMapp is the netaffx database from Affymetrix.

Detailed Description of the Invention

Methods of Monitoring Disease or Condition

In the present invention, it will be understood that the determination of cells "obtained from the site" of a disease or condition in a patient is reference to an *in vitro* method practiced on a sample after removal from the body. The removal of the body sample, e.g. in a biopsy, is not part of the invention as such.

As explained above, our methodology used to identify the genes

of Table 1, 2 or 3 is a useful means for monitoring the progression of disease or other conditions associated with pterygium. In the treatment of diseases or conditions associated with pterygium, the clinician will look for a
5 response in which the genes show transcript levels uncharacteristic of normally regulated tissue.

The up or down-regulation of the genes we have identified can be made during a course of treatment of a patient so that the
10 effectiveness of the treatment can be gauged. For example, many cancer treatments rely upon a cocktail of different anti-cancer agents. The effectiveness of any one particular cocktail may differ from patient to patient, or during the course of treatment in the patient where cells become
15 resistant to one or more of the drugs. In an analogous manner, the response to treatment of pterygia in a patient can be monitored and changes in the response to treatment in that patient can be used to gauge the effectiveness of the treatment or to alter the course of treatment.

20 In this aspect of the invention, the comparison can be made with the transcript levels obtained from the same site in the patient at the same time. Control tissue is available from the unexposed superior bulbar conjunctiva of the same eye from
25 which the pterygial tissue is removed and can be removed from the patient at the same time as the diseased tissue. Alternatively, the comparison may be made with transcript levels of pterygial cells taken from the same patient over a period of time for example where the disease is recurrent or
30 where non-surgical therapy is used in the treatment of the condition. Another option is to provide a control baseline sample or historical record from another patient, or, more preferably, a population of patients. Preferably, the control

cells are conjunctival cells from the same patient.

In a preferred aspect, the invention is performed by looking at the transcript pattern of a plurality of genes. This is because we have found that in individual subjects, the transcript level of individual genes may vary. It is therefore desirable that the transcript level is assessed for several genes. For example, the genes assessed could include those responsible for cell growth and maintenance, cell communication, cell adhesion, cyclins, translation initiation factors, signal transduction, oncogenes, extra-cellular matrix, UV sensitive and cancer-related genes.

We have discovered that a particularly preferred marker for pterygium are the collagen genes, particularly collagen I α 1 and collagen III α 1. Further preferred markers include the β -microseminoprotein, tyrosinase-related protein 1 and matrix Gla protein.

Generally, the transcript level of at least 5, preferably at least 10 and more preferably at least 20 genes is determined. The genes are desirably from at least two or more of the above categories, for example from at least 3, 4 or 5 of the above categories.

The transcript level of a gene or genes may be determined by any suitable means. Where many different gene transcripts are being examined, a convenient method is by hybridization of the sample (either directly or after generation of cRNA or cDNA) to a gene chip array.

Where gene chip technology is used, the genes or clones (this term used herein includes ESTs) are all commercially

available, e.g. from Affymetrix or from Incyte Pharmaceuticals Inc. The clones may be used to assemble appropriate chips or microarrays, as is known in the art. Generally, methods for the provision of microarrays and their use may also be found
5 in, for example, WO84/01031, WO88/1058, WO89/01157, WO93/8472, WO95/18376/ WO95/18377, WO95/24649 and EP-A-0373203 and reference may also be made to this and other literature in the art.

10 Table 1 provides the names of genes and these may be used to obtain their DNA sequences from databases such as Genbank. In addition, the particular sequences on the Incyte chip we have used may be determined by the Incyte reference number supplied in Table 1, either as the Incyte clone ID or as the Incyte
15 Genbank ID, and these may be related directly to Genbank reference numbers. The EST gene sequences are also given Incyte reference numbers. Those of skill in the art may refer to either of the Incyte reference numbers or the appropriate Genbank reference number in practicing the present invention.

20 In a similar manner, Tables 2 and 3 provide the names, location and ID numbers for genes determined using the Affymetrix chip. These tables also provide details of the function of the genes and the pathway in which these genes, or
25 the protein products thereof, are involved.

Alternatively, or in addition, quantitative PCR methods may be used, e.g. based upon the ABI TaqMan™ technology, which is widely used in the art. It is described in a number of prior
30 art publications, for example reference may be made to WO00/05409. PCR methods require a primer pair which target opposite strands of the target gene at a suitable distance apart (typically 50 to 300 bases). Suitable target sequences

for the primers may be determined by reference to Genbank sequences as mentioned above.

Gene Chips

5 Although the prior art provides a gene chip which includes, as part of a very large array, the genes of Table 1, 2 or 3, the identification of a relatively small set of genes of diagnostic and prognostic use in the present situation allows the provision of a small chip specifically designed to be
10 suitable for use in the present invention.

Thus the invention provides a gene chip array comprising at least one nucleic acid suitable for detection of at least one gene shown in Table 1, 2 or 3; optionally a control specific
15 for said at least one gene; and optionally at least one control for said gene chip. Desirably, the number of sequences in the array will be such that where the number of nucleic acids suitable for detection of the Table 1, 2 or 3 transcripts is n , the number of control nucleic acids specific
20 for individual transcripts is n' , where n' is from 0 to $2n$, and the number of control nucleic acids (e.g. for detection of "housekeeping" transcripts or other pterygial transcripts) on said gene chip is m where m is from 0 to 100, preferably from 1 to 30, then $n + n' + m$ represent at least 50%, preferably
25 75% and more preferably at least 90% of the nucleic acids on said chip.

As described above, the chip of the present invention is designed specifically to be suitable for use in the present
30 invention. It is therefore preferred that the chip has from 40 to 8000 nucleic acids, more preferably from 50 to 400 nucleic acids, and most preferably from 50 to 100 nucleic acids. On such a chip, there will preferably be nucleic acids

specific for detection of at least 10, for example at least 15, preferably at least 20, more preferably at least 30, more preferably 35 and most preferably at least 40 of the transcripts of Table 1, 2 3.

5

Assay Methods

The assay method of the present invention may be practiced in a wide variety of formats, for example on protein or nucleic acid components or in whole cells in culture. The whole cells
10 may be in the form of primary tissue taken from a patient with pterygia or may be a culture of cells.

One assay comprises:

- 15 (a) providing a protein encoded by a transcript of Table 1, 2 or 3;
- (b) bringing said protein into contact with a candidate modulator of its activity; and
- (c) determining whether said candidate modulator is capable of modulating the activity of said protein.

20

In this assay method, the determination of modulation of activity will depend upon the nature of the protein being assayed. For example, proteins with enzymatic function may be assayed in the presence of a substrate for the enzyme, such
25 that the presence of a modulator capable of modulating the activity results in a faster or slower turnover of substrate. The substrate may be the natural substrate for the enzyme or a synthetic analogue. In either case, the substrate may be labelled with a detectable label to monitor its conversion
30 into a final product.

The assay method will be tailored to the nature of the protein being assayed. For example, we have identified a number of

transcription factors. Transcription factors orchestrate the response of a cell to a given stimulus, controlling multiple genes necessary for the cell to grow, divide, migrate, differentiate etc. The transcription factors may themselves
5 be regulated (in a cascade like fashion). Hence these genes listed here are of key importance. The activity of such genes can thus be assayed by looking at transcriptional responses, or by looking at DNA binding activity.

10 For these and other proteins with DNA binding activity, such as transcription regulators, the DNA binding or transcriptional activating activity may be determined, wherein a modulator is able to either enhance or reduce such activity. For example, DNA binding may be determined in a mobility shift
15 assay. Alternatively, the DNA region to which the protein binds may be operably linked to a reporter gene (and additionally, if needed, a promoter region and/or transcription initiation region between said DNA region and reporter gene), such that transcription of the gene is
20 determined and the modulation of this transcription, when it occurs, can be seen. Suitable reporter genes include, for example, chloramphenicol acetyl transferase or more preferably, fluorescent reporter genes such as green fluorescent protein.

25 Another group of proteins are the intracellular signal transduction proteins, such as kinases, growth factors and proteases/peptidase. Kinases frequently are involved in the transduction of extra-cellular signals to the inside of the
30 cell. As such they are essential for cell to cell communication and are essential for the coordinated response of a tissue. Some of the effects of steroid hormones can be replicated by administration of specific growth factors and

blocked by inhibiting growth factor receptor activation. This indicates that the growth factors and their receptors are necessary for the correct function of the conjunctiva. Thus the identification of the steroid responsive growth factor and/or receptors opens new avenues for therapeutic intervention.

Phosphatases also play an essential role in this as they turn "off" the signal from the kinases. Proteases (and peptidases) play at least 2 roles in this respect. They are able to activate growth factors (by cleavage of a precursor, or by releasing bound forms from the extracellular matrix), or inactivate them (by degradation); they also digest the extracellular matrix which is essential for cell migration, differentiation and tissue remodelling. Assay formats designed to examine such proteins may involve the use of labelled phosphate groups and a suitable substrate.

We have also identified a few cell death and cell survival factors. The structural and functional integrity of a tissue is maintained by the balance of cell proliferation and cell death within it. In a tissue such as conjunctiva, the rate of these processes will vary over time and be tightly regulated. Thus, the genes involved in these processes (or their regulation) have far reaching consequences for the correct function of the conjunctiva. Such proteins may be assayed for modulators which bind directly to them in a manner which interferes with their function, such as binding to a target protein involved in progressing the course of cell death.

Other classes include signalling factors, and proteins with a ligand binding function, such as receptors, such as S100. The candidate modulator may be examined for ligand binding

function in a manner that leads to antagonism or agonism of the ligand binding property.

Candidate modulator compounds may be natural or synthetic chemical compounds used in drug screening programmes.

Extracts of plants, microbes or other organisms, which contain several characterised or uncharacterised components may also be used. Combinatorial library technology (including solid phase synthesis and parallel synthesis methodologies) provides an efficient way of testing a potentially vast number of different substances for ability to modulate an interaction. Such libraries and their use are known in the art, for all manner of natural products, small molecules and peptides, among others. Many such libraries are commercially available and sold for drug screening programmes of the type now envisaged by the present invention.

A further class of candidate modulators are antibodies or binding fragment thereof which bind a protein target.

Example antibody fragments, capable of binding an antigen or other binding partner are the Fab fragment consisting of the VL, VH, Cl and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an antibody; the dAb fragment which consists of a VH domain; isolated CDR regions and F(ab')₂ fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included. An antibody specific for a protein may be obtained from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on

their surfaces; for instance see WO92/01047. Such a technique allows the rapid production of antibodies against an antigen, and these antibodies may then be screened in accordance with the invention.

5

Another class of candidate molecules are peptides based upon a fragment of the protein sequence to be inhibited. In particular, fragments of the protein corresponding to portions of the protein which interact with other proteins or with DNA may be a target for small peptides which act as competitive inhibitors of protein function. Such peptides may be for example from 5 to 20 amino acids in length.

The peptides may also provide the basis for design of mimetics. Such mimetics will be based upon analysis of the peptide to determine the amino acid residues or portions of their side chains essential and important for biological activity to define a pharmacophore followed by modelling of the pharmacophore to design mimetics which retain the essential residues or portions thereof in an appropriate three-dimensional relationship. Various computer-aided techniques exist in the art in order to facilitate the design of such mimetics.

Cell based assay methods can be configured to determine expression of the gene either at the level of transcription or at the level of translation. Where transcripts are to be measured, then this may be determined using the methods of the first aspect of the invention described above, e.g. on gene chips, by multiplex PCR, or the like.

Cell based assay methods may be used to screen candidate modulators as described above. They may also be used to

screen further classes of candidate modulator, including antisense oligonucleotides. Such oligonucleotides are typically from 12 to 25, e.g. about 15 to 20 nucleotides in length, and may include or consist of modified backbone
5 structures, e.g. methylphosphonate and phosphorothioate backbones, to help stabilise the oligonucleotide. The antisense oligonucleotides may be derived from the coding region of a target gene or be from the 5' or 3' untranslated region. Candidate molecules may further include RNAi, i.e.
10 short double stranded RNA molecules which are sequence specific for a gene transcript.

Modulators obtained in accordance with the present invention may be used in methods of modulating conjunctival (pterygial)
15 tissue growth in a human patient. Generally, the modulator will be formulated with one or more pharmaceutically acceptable carriers suitable for a chosen route of administration to a subject. For solid compositions, conventional non-toxic solid carriers include, for example,
20 pharmaceutical grades of mannitol, lactose, cellulose, cellulose derivatives, starch, magnesium stearate, sodium saccharin, talcum, glucose, sucrose, magnesium carbonate, and the like may be used. Liquid pharmaceutically administrable compositions can, for example, be prepared by dissolving,
25 dispersing, etc, a modulator and optional pharmaceutical adjuvants in a carrier, such as, for example, water, saline aqueous dextrose, glycerol, ethanol, and the like, to thereby form a solution or suspension. If desired, the pharmaceutical composition to be administered may also contain minor amounts
30 of non-toxic auxiliary substances such as wetting or emulsifying agents, pH buffering agents and the like, for example, sodium acetate, sorbitan monolaurate, triethanolamine sodium acetate, sorbitan monolaurate, triethanolamine oleate,

etc. Actual methods of preparing such dosage forms are known, or will be apparent, to those skilled in this art; for example, see Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pennsylvania, 15th Edition, 1975.

5 The composition or formulation to be administered will, in any event, contain a quantity of the active compound(s) in an amount effective to alleviate the symptoms of the subject being treated.

10 Routes of administration may depend upon the precise condition being treated.

Peptides of the invention

We have identified a number of ESTs which have not previously
15 been associated with pterygial growth. The identification of these ESTs will allow us to extend these ESTs and thus identify nucleic acid sequences and where applicable the amino acid sequences they encode, which for the first time can be utilised in a useful manner as markers of pterygium and
20 targets in the development of modulators of such pterygial tissue growth.

Thus in a further aspect, the invention provides the use of an EST as identified in Table 1 in the identification of nucleic
25 acid sequences suitable for use as a marker of pterygium. The present invention also covers the amino acid sequences encoded by the nucleic acid sequences mentioned herein or a fragment of such a sequence having at least 15 amino acids. The invention also provides peptides which have at least 50%
30 sequence identity to said peptides or a fragment thereof.

As used herein, "peptide" refers to any chain of naturally occurring amino acids which are encoded by the mammalian

genetic code of 15 or more amino acids in length. The term thus includes full length proteins as well as fragments thereof.

- 5 An isolated peptide of the invention will be free or substantially free of material with which it is naturally associated such as other peptides with which it is found in the cell. The peptides may of course be formulated with diluents or adjuvants and still for practical purposes be
- 10 isolated - for example the peptides will normally be mixed with gelatin or other carriers if used to coat microtitre plates for use in immunoassays. The peptides may be glycosylated, either naturally or by systems of heterologous eukaryotic cells, or they may be (for example if produced by
- 15 expression in a prokaryotic cell) unglycosylated. Peptides may phosphorylated and/or acetylated.

A peptide of the invention may also be in a substantially purified form, in which case it will generally comprise the

20 peptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the peptide in the preparation is a peptide of the invention.

Peptides of the invention may be modified for example by the

25 addition of histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell or uptake into a cell.

Peptides which are amino acid sequence variants, alleles,

30 derivatives or mutants are also provided by the present invention, such forms having at least 50% sequence identity, for example at least 70%, 80%, 90%, 95%, 98% or 99% sequence identity to one of the peptides of the present invention. A

peptide which is a variant, allele, derivative or mutant may have an amino acid sequence which differs from that given in the listing by one or more of addition, substitution, deletion and insertion of one or more (such as from 1 to 20, for
5 example 2, 3, 4, or 5 to 10) amino acids.

The percentage homology (also referred to as identity) of DNA and amino acid sequences can be calculated using commercially available algorithms. The following programs (provided by the
10 National Center for Biotechnology Information) may be used to determine homologies: BLAST, gapped BLAST, BLASTN and PSI-BLAST, which may be used with default parameters.

A "fragment" means a stretch of amino acid residues of at
15 least about fifteen amino acids, preferably at least about 20 to 30 or more contiguous amino acids. Fragments of the peptides comprise epitopes useful for raising antibodies to a portion of the amino acid sequences of the sequences set out herein below. Preferred epitopes are those to which antibodies
20 are able to bind specifically.

A peptide according to the present invention may be isolated and/or purified (e.g. using an antibody) for instance after production by expression from encoding nucleic acid. Peptides
25 according to the present invention may also be generated wholly or partly by chemical synthesis, for example in a step-wise manner. The isolated and/or purified peptide may be used in formulation of a composition, which may include at least one additional component, for example a pharmaceutical
30 composition including a pharmaceutically acceptable excipient, vehicle or carrier. A composition including a peptide according to the invention may be used in prophylactic and/or therapeutic treatment.

A peptide according to the present invention may be used as an immunogen or otherwise in obtaining specific antibodies. Antibodies are useful in purification and other manipulation of peptides, diagnostic screening and therapeutic contexts.

5

A peptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic (possibly including prophylactic) context.

10

A peptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the peptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I , enzymes, antibodies, polynucleotides

15

and linkers such as biotin. Labelled peptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a peptide of the invention in a sample. Peptides or labelled peptides of the invention may also be used in serological or cell mediated immune assays for the detection of immune reactivity to said peptides in animals and humans using standard protocols.

20

A peptide or labelled peptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well or dipstick.

25

Such labelled and/or immobilized peptides may be packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

30

Such peptides and kits may be used in methods of detection of antibodies to such peptides present in a sample or active portions or fragments thereof by immunoassay.

Immunoassay methods are well known in the art and will generally comprise:

- 5 (a) providing a peptide comprising an epitope bindable by an antibody against said peptide;
- (b) incubating a biological sample with said peptide under conditions which allow for the formation of an antibody-antigen complex; and
- 10 (c) determining whether antibody-antigen complex comprising said peptide is formed.

The identification of the peptide expressed by the genes identified herein enables assays to be developed to identify further cellular proteins with which the peptide is
15 associated. For example, peptides of the present invention may be required in a regulatory pathway in which their function is to interact with other factors which in turn promote or maintain essential cellular functions associated with cell cycle control. The peptides of the present
20 invention may be used in two-hybrid assays to determine cellular factors with which they become associated.

Two-hybrid assays may be in accordance with those disclosed by Fields and Song, 1989, *Nature* 340; 245-246. In such an assay
25 the DNA binding domain (DBD) and the transcriptional activation domain (TAD) of the yeast GAL4 transcription factor are fused to the first and second molecules respectively whose interaction is to be investigated. A functional GAL4 transcription factor is restored only when two molecules of
30 interest interact. Thus, interaction of the molecules may be measured by the use of a reporter gene operably linked to a GAL4 DNA binding site which is capable of activating transcription of said reporter gene. Other transcriptional

activator domains may be used in place of the GAL4 TAD, for example the viral VP16 activation domain. In general, fusion proteins comprising DNA binding domains and activation domains may be made.

5

In the present case peptides of the invention may be expressed as fusion proteins with an appropriate domain and candidate second peptides with which those of the invention might associate can be produced as fusion proteins with an appropriate corresponding domain. Alternatively libraries such as phage display libraries of such fusion proteins may be screened with a fusion peptide of the invention.

The provision of the novel peptides enables for the first time the production of antibodies able to bind them specifically. Such an antibody may be specific in the sense of being able to distinguish between the peptide it is able to bind and other peptides of the same species for which it has no or substantially no binding affinity (e.g. a binding affinity of at least about 1000x worse). Specific antibodies bind an epitope on the molecule which is either not present or is not accessible on other molecules. Antibodies according to the present invention may be specific for the wild-type peptide. Antibodies according to the invention may be specific for a particular mutant, variant, allele or derivative peptide as between that molecule and the wild-type peptide, so as to be useful in diagnostic and prognostic methods as discussed below. Antibodies are also useful in purifying the peptides to which they bind, e.g. following production by recombinant expression from encoding nucleic acid.

30

Preferred antibodies according to the invention are isolated, in the sense of being free from contaminants such as

antibodies able to bind other peptides and/or free of serum components. Monoclonal antibodies are preferred for some purposes, though polyclonal antibodies are within the scope of the present invention. Monoclonal antibodies obtained from
5 non-human animals may be humanised using CDR grafting techniques.

Antibodies may be obtained using techniques which are standard in the art. Methods of producing antibodies include
10 immunising a mammal (e.g. mouse, rat, rabbit) with a peptide of the invention. Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or
15 immunoprecipitation may be used (Armitage et al, Nature, 357:80-82, 1992).

As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a protein may be obtained
20 from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047.

25 The term "antibody" should be construed as covering any binding substance having a binding domain with the required specificity. Example antibody fragments, capable of binding an antigen or other binding partner are the Fab fragment
30 consisting of the VL, VH, C1 and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an antibody; the dAb fragment which consists of a VH domain;

isolated CDR regions and F(ab')₂ fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

5

A hybridoma producing a monoclonal antibody according to the present invention may be subject to genetic mutation or other changes.

10 The reactivities of antibodies on a sample may be determined by any appropriate means. Tagging with individual reporter molecules is one possibility. The reporter molecules may directly or indirectly generate detectable, and preferably measurable, signals. The linkage of reporter molecules may be
15 directly or indirectly, covalently, e.g. via a peptide bond or non-covalently. Linkage via a peptide bond may be as a result of recombinant expression of a gene fusion encoding antibody and reporter molecule.

20 Reporters include macromolecular colloidal particles or particulate material such as latex beads that are coloured, magnetic or paramagnetic, and biologically or chemically active agents that can directly or indirectly cause detectable signals to be visually observed, electronically detected or
25 otherwise recorded. These molecules may be enzymes which catalyse reactions that develop or change colours or cause changes in electrical properties, for example. They may be molecularly excitable, such that electronic transitions between energy states result in characteristic spectral
30 absorptions or emissions.

Antibodies may modulate the activity of the peptide to which they bind and so, if that peptide has a deleterious effect in

an individual, may be useful in a therapeutic context (which may include prophylaxis).

Nucleic Acids

- 5 The invention also provides an isolated nucleic acid which:
- (a) encodes a peptide of the invention, including a nucleic acid which can be identified using the ESTs set out in Table 1; or
 - (b) comprises or consists essentially of any of an extended
- 10 EST sequence which has not been translated.

The invention also provides a nucleic acid which has at least 70% homology to (a) and (b) above. Preferably the degree of sequence homology is at least 80%, e.g. at least 90%,

15 preferably at least 95%, more preferably at least 98% and most preferably at least 99%.

The nucleic acids may be in the form of a vector, such as an expression vector wherein said nucleic acid is operably linked

20 to a promoter heterologous to said nucleic acid. The promoter will be compatible with a desired host cell, and such host cells form a further aspect of the invention.

Nucleic acids encoding or associated with these new sequences

25 may be used in methods of detecting the presence or absence of said gene in a human subject, said method comprising;

- (a) bringing a sample of nucleic acid from said subject into contact, under hybridizing conditions, with a nucleic acid of the invention; and
- 30 (b) determining whether said nucleic acid of the invention has been able to hybridize to a homologous sequence in said sample nucleic acid.

Nucleic acid includes DNA (including both genomic and cDNA) and RNA, and also synthetic nucleic acids, such as those with modified backbone structures intended to improve stability of the nucleic acid in a cell. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or lifespan of polynucleotides of the invention. Where nucleic acid according to the invention includes RNA, reference to the sequences shown in the accompanying listings should be construed as reference to the RNA equivalent, with U substituted for T.

Nucleic acid of the invention may be single or double stranded polynucleotides. Single stranded nucleic acids of the invention include anti-sense nucleic acids.

Generally, nucleic acid according to the present invention is provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally associated, such as free or substantially free of nucleic acid flanking the gene in the human genome, except possibly one or more regulatory sequence(s) for expression.

Preferably the degree of sequence identity in either case is at least 80%, such as at least 90%, 95%, 98% or 99%.

The invention also provides nucleic acids which are fragments of the nucleic acids described herein. Particular nucleic

acids which are preferred include nucleic acids which consist essentially of from 15 to 50, for example from 15 to 35, 18 to 35, 15 to 24, 18 to 30, 18 to 21 or 21 to 24 nucleotides of a sequence having at least 70% homology to the nucleic acid sequences of the present invention.

The term "consist essentially of" refers to nucleic acids which do not include any additional 5' or 3' nucleic acid sequences. In a further aspect of the invention, nucleic acids of the invention which consist essentially of from 15 to 30 nucleotides as defined above may however be linked at the 3' but preferably 5' end to short (e.g. from 4 to 15, such as from 4 to 10 nucleotides) additional sequences to which they are not naturally linked. Such additional sequences are preferably linkers which comprise a restriction enzyme recognition site to facilitate cloning when the nucleic acid of the invention is used for example as a PCR primer.

Nucleic acids of the invention, particularly short (less than 50) sequences useful as probes and primers may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S , fluorescent labels, enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides or primers of the invention and may be detected using by techniques known per se.

Also included within the scope of the invention are antisense sequences based on the nucleic acid sequences described herein, preferably in the form of oligonucleotides, particularly stabilized oligonucleotides, or ribozymes. Antisense oligonucleotides may be designed to hybridise to the complementary sequence of nucleic acid, pre-mRNA or mature mRNA, interfering with the production of polypeptide encoded

by a given DNA sequence, so that its expression is reduced or prevented altogether. In addition to the coding sequence, antisense techniques can be used to target the control sequences of genes, e.g. in their 5' flanking regions. The construction of antisense sequences and their use is described in Peyman and Ulman, Chemical Reviews, 90:543-584, (1990), Crooke, Ann. Rev. Pharmacol. Toxicol., 32:329-376, (1992), and Zamecnik and Stephenson, P.N.A.S, 75:280-284, (1974).

Nucleic acid sequences encoding all or part of the sequences of the invention can be readily prepared by the skilled person using the information and references contained herein and techniques known in the art, including, for example the polymerase chain reaction (PCR) to amplify samples of such nucleic acid, e.g. from genomic sources, (ii) chemical synthesis, or (iii) preparing cDNA sequences by standard cloning methodology. Such techniques may be used to obtain all or part of the sequences described herein.

Nucleic acid according to the present invention, such as a full-length coding sequence or oligonucleotide probe or primer, may be provided as part of a kit, e.g. in a suitable container such as a vial in which the contents are protected from the external environment. The kit may include instructions for use of the nucleic acid, e.g. in PCR and/or a method for determining the presence of nucleic acid of interest in a test sample. A kit wherein the nucleic acid is intended for use in PCR may include one or more other reagents required for the reaction, such as polymerase, nucleosides, buffer solution etc. The nucleic acid may be labelled. A kit for use in determining the presence or absence of nucleic acid of interest may include one or more articles and/or reagents for performance of the method, such as means for providing the

test sample itself, e.g. a swab for removing cells from the buccal cavity or a syringe for removing a blood sample (such components generally being sterile). In a further aspect, the present invention provides an apparatus for screening nucleic acid, the apparatus comprising storage means including the nucleic acid of the invention or fragment thereof, the stored sequence being used to compare the sequence of the test nucleic acid to determine the presence of mutations.

10 Nucleic acids of the invention are thus useful in screening a test sample containing nucleic acid for the presence of alleles, mutants and variants, which may be of further clinical relevance in diagnosis or prognosis.

15 **Vectors**

The identification of a number of ESTs associated with pterygial tissue growth provides the basis for novel vector systems useful in the aspects of the invention described above, as well as further aspects described herein below.

20 Thus, expression vectors for the expression of proteins of the invention for the expression of RNA in a sense or antisense orientation of DNA sequences of the invention form a further aspect of the invention.

25 Preferably, a nucleic acid sequence of the invention is in a vector operably linked to a control sequence which is capable of providing for the expression of the sequence by a host cell, i.e. the vector is an expression vector.

30 The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a

way that expression of the coding sequence is achieved under condition compatible with the control sequences.

Suitable host cells include bacteria, eukaryotic cells such as mammalian and yeast, and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous peptide include Chinese hamster ovary cells, HeLa cells, baby hamster kidney cells, COS cells and many others.

The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid, nucleic acid sequences so that where a peptide is produced it is in the form of a fusion with secretion signals so that the peptide produced in the host cell is secreted from the cell.

The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector.

Vectors may further include enhancer sequences, terminator fragments, polyadenylation sequences and other sequences as appropriate.

Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector may also be adapted to be used *in vivo*, for example in methods of gene therapy. Systems for cloning and expression of a peptide in a variety of different host cells are well known. Vectors include gene therapy vectors, for example vectors based on adenovirus, adeno-associated virus, retrovirus (such as HIV or MLV) or alpha virus vectors.

Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. For example, yeast promoters
5 include *S. cerevisiae* GAL4 and ADH promoters, *S. pombe* nmt1 and adh promoter. Mammalian promoters include the metallothionein promoter which is can be included in response to heavy metals such as cadmium. Viral promoters such as the SV40 large T antigen promoter or adenovirus promoters may also
10 be used. All these promoters are readily available in the art.

Vectors for production of peptides of the invention of for use in gene therapy include vectors which carry a mini-gene
15 sequence.

Vectors may be transformed into a suitable host cell as described above to provide for expression of a peptide of the invention. Thus, in a further aspect the invention provides a
20 process for preparing peptides according to the invention which comprises cultivating a host cell transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the peptides, and recovering the expressed
25 peptides. Peptides may also be expressed using in vitro systems, such as reticulocyte lysate.

The invention is illustrated by the following examples.

Example 1

5 **Methods:**

Primary pterygium tissue was obtained from 29 patients undergoing pterygium excision with conjunctiva autografting, utilizing remnant superior bulbar conjunctiva adjacent to the graft harvest site as control tissue. Remnants of the
10 superior conjunctiva and pterygia tissue were snap frozen for each patient. For the purpose of mRNA extraction, each type of tissue sample was pooled. Total RNA was processed for analysis by Incyte using the human UniGem V 2.33 array that contained 9,128 annotated genes or EST clusters. The results
15 were analyzed by GemTools and Gene Spring software.

Results:

Thirty nine genes were upregulated about 2 fold or more in the pterygium tissue when compared with the control conjunctiva.
20 The array passed all quality control checks and the sensitivity was 1 in 100,000. Upregulated genes were found in a number of major categories: Cell growth and maintenance, cell communication, cell adhesion, cyclins, translation initiation factors, signal transduction, oncogenes and cancer-
25 related. Collagen is major structural protein of pterygia tissue. The gene for collagen type I, alpha (I) was upregulated, collagen type I, alpha 1 increased about 2 fold compared to the control conjunctiva. Few changes in transcription were found for common growth factors or genes
30 for apoptosis.

Discussion:

We have carried out the first DNA array study of pterygium

using tissue from 29 patients with primary pterygium. Patients were obtained over several months and no special criteria were employed except that they were confirmed as primary pterygia. Due to the demand for large amounts of aRNA for the Incyte
5 array at that time, we pooled all of the samples. All tissues were rapidly frozen in liquid nitrogen at the time of removal, and maintained until pooling for extraction of total message. The Incyte Human UniGem V 2.33 array was used, total message was processed by Incyte and the array was completed by Incyte.
10 The array contained 8632 entries of which 8502 were genes and 130 were ESTs, or fragments of genes. Each probe averaged 1000 base pairs.

Quality control issues of array studies are also important and
15 Incyte had a series of controls that were carried out and determined to be all within limits. Further inspection showed that about 39 genes in several categories were upregulated. Additionally there were 3 ESTs upregulated. These included five extra-cellular matrix, two cytoskeletal proteins, three
20 cell communication/adhesion proteins, two oncogenes and five genes involved in intracellular signal transduction pathways. Collagen alpha 1(I) was strongly upregulated, a finding that is corroborated by electron microscopic analysis of pterygium tissue. Pathways involving these gene products may be useful
25 in the future as potential therapeutic targets.

The overall sensitivity or detection level of the occurrence of a particular RNA species was determined to be 1:100,000, that is one copy in 100,000. The level of significance of an
30 upregulated gene was determined by Incyte to be greater than 1.74. In other words when the balanced differential expression was calculated which is essentially the ratio between the expression level of a gene expressed in control and disease,

the log change of the disease expression compared to control has to be greater than 1.74.

The results indicated that the upregulated genes occupied several functional classes. These classes are shown in Table 4, below. These genes characterize the abnormal phenotype of pterygium and can have important roles in the transformation of the tissue to the disease state through production of their associated proteins. Not all of the genes that were found to be upregulated are necessarily targets for therapy. Thus, the genes listed below are considered to be likely candidates for drug development; however, other genes listed may also be considered.

Table 4: Classes of Gene Function

<u>Class of Gene</u>	<u>Gene</u>
<i><u>Extra-cellular Matrix</u></i>	Fibronectin 1 Collagen type I Lumican Matrix Gla protein Extracellular matrix protein 2, female organ and adipocyte specific
<i><u>Oncogene</u></i>	DEK
<i><u>Genes whose regulation is affected by UVB</u></i>	Fibronectin 1 Collagen I, Alpha I P311 Tyrosinase-related protein
<i><u>Genes related to proliferation</u></i>	Mitogen-activated protein kinase kinase 7
<i><u>Cell communication and junctions</u></i>	Desmoglein Novel RGD-containing protein Multimerin
<i><u>Cell growth and maintenance</u></i>	ATPase, Ca ++ transporting cardiac muscle, slow twitch 2
<i><u>Intracellular signalling</u></i>	Cystatin A

Example 2**Methods:**

One group of 8 primary pterygium samples (age range: 40-60 years) and another group of 3 control conjunctiva tissues (5 pooled tissues in each) were included. At the time of surgery and after removal, all tissues were rapidly frozen in liquid nitrogen. Total RNA was extracted and prepared for Affymetrix arrays according to Affymetrix standard protocol. Each of the samples was hybridized onto Human Genome Chip U133A. The chips were scanned 2X and the expression level for each probe set were scaled to 500. The data analysis and visualization have been performed using Statistica 6.0, Microarray Suite v.5.0, Data Mining Tool v.0.3 and GeneSpring v.5.

Results:

In 22,283 total probe sets, 8,673-11,225 (38.9%-50.4%) genes and EST's were present in pterygium and 9,876-10,907 (44.3%-48.9%) genes and EST's were present in conjunctiva. The number of genes exhibiting marginal expression for both groups was as low as 1.6-1.9%. Histogram plots of the signal value showed that the data had a non-normal distribution. After applying the Mann-Whitney test with cut-off of $p=0.05$ to determine the significance and the direction of the 'Change Call', 979 entries for genes that were up-regulated and 934 for genes that were down-regulated remained. Of these genes, over 500 in each category were determined to be absent and were omitted from further consideration (see Table 5). Several hundred pterygium genes showed an increase or decrease more than 2 fold ($SLR \geq 1$ or ≤ -1) in change compared to the control conjunctiva (Consistency of change is 75% from 24

comparison pairs). The combined result from the tests gave 28 genes that were up-regulated and 21 genes down-regulated.

K-Mean clustering divided genes in both groups, up regulated
5 and down regulated, into subgroups based on their expression
patterns. This produced groups of genes with a high degree of
similarity within each group. The up regulated genes in this
case, showed more distinctive results than the down regulated
genes. Up regulated genes were found in several bioprocess
10 categories and many were involved in inflammatory response
pathways. Down regulated genes were involved in apoptosis,
MAPK Cascade and in G-Protein and TGF-Beta cell signalling
pathways.

15 General Affymetrix procedures and good examples of the use of
K-means clustering are described in, for example, Bowcock AM
et al., Insights into psoriasis and other inflammatory disease
from large-scale gene expression studies, Human Molecular
Genetics (2001) **17**:1793-1805.

Table 5: Application of Mann-Whitney test with $p=0.05$

	INCREASE		DECREASE	
Detection Call	Absent-Present And Present-Present	8280	Present-Absent And Present-Present	8281
Mann-Whitney Test α 0.05	$p\text{-value} \leq 0.05$ Category: Up	439	$p\text{-value} \geq 0.05$ Category: Down	470
Change Direction	Change Call:increase 75% across all 24 comparison	74	Change Call:decrease 75% across all 24 comparison	90
Signal Log Ratio (SLR)	$SLR \geq 1$ 75% across all 24 comparison	38	$SLR \leq 1$ 75% across all 24 comparison	25
Combined result		28		21

Discussion:

- 5 In this second example, the up or down regulation of genes has been determined using an Affymetrix oligonucleotide array.

Comparing the data sets of Examples 1 and 2, it is clear that there are many similarities in the genes that are highly up regulated. It is not possible to compare these two data sets statistically as there are fundamental differences in methodology. However, the two data sets are amenable to specific types of data analysis and they have in common procedures that allow decisions to be made as to which messages are the most highly expressed. Thus, some

conclusions can be made about the level of expression. The advantage of the Affymetrix data is that it included 8 separate primary pterygia and 3 groups of five each pooled normal conjunctiva. This allowed appropriate statistical analysis to be carried out to arrive at the most highly regulated genes in this disease.

For example, in the data set of Example 1 we found that fibronectin I is highly expressed, and that is still the case after adding additional data from Example 2. In addition, beta microseminoprotein remains a highly expressed gene throughout the two data sets. A number of collagen genes such as collagen I, alpha 1 is highly expressed in both data sets. In addition, tyrosinase-related protein 1, and matrix Gla protein both remain up regulated. The collagen genes are well represented in the Affymetrix array and we now have evidence for abnormal collagen, formed in the pterygium, collagen type III, alpha 1, which is found in Ehles-Danlos syndrome which is characterized by collagen without structural integrity. These additional data affirm our conclusion that pterygium is an inflammatory disease associated with defective tissue. The UV response factor remains in both data sets with the consistency of the finding of the up regulation of serine and cysteine proteases.

Thus, the data sets from the two Examples provide strong evidence for the assertion that new genes have been revealed that for the first time characterize human primary pterygium and these genes may form the basis of tests or therapies to diagnose or to treat the disease.

TABLE 1

GeneID (Incyte)	GeneName	Diff Expr	Genebank Access Num	Locus	Incyte CloneID	GenBank Id (Incyte)
-128148	fibronectin 1	4.6	AW385690	2q34	3553729	pINCY
-177934	osteoblast specific factor 2 (fasciclin I-like)	3.6	AU117842	13	1994715	pSport1
-119389	deiodinase, iodothyronine, type II	3.1	AF007144	14q24.2- q24.3	2714959	pINCY
-124642	desmoglein 1	3	AF097935	18q12.1	3561805	pINCY
-120057	lumican	2.7	AU120442	12q21.3-q22	1228124	pSport1
-127808	hypothetical protein FLJ20277	2.7	AI922083		2006061	pBlue
-177391	P311 protein	2.6	AF119859		1555545	pINCY
-121697	tyrosinase-related protein 1	2.3	X51420	9p23	467240	pBlue
-120611	antigen identified by monoclonal antibody Ki-67	2.2	X65550	10q25-qter	2470485	pINCY
-123537	microseminoprotein, beta-	2.3	AA574208	10q11.2	3970665	pINCY
-121768	cystatin A (stefin A)	2.2	AI680589	3q21	1821453	pINCY
-178467	Homo sapiens clones 24714 and 24715 mRNA sequence	2.1	R56641		2599229	pINCY
-178677	extracellular matrix protein 2, female organ and adipocyte specific	2.1	NM_001393	9q22.3	2886607	pINCY
-119113	nuclear receptor subfamily 3, group C, member 2	2.1	M16801	4q31.1	1678071	pINCY
-123670	collagen, type I, alpha 1	2	AU117762	17q21.3-q22	782235	pSport1
-128423	EST	2	AA219230		891949	pSport1
-120526	selectin P (granule membrane protein 140kD, antigen CD62)	2	M25322	1q22-q25	3688202	pINCY
-128604	ESTs	2	AW129530		2293824	pSport1

GeneID (Incyte)	GeneName	Diff Expr	Genebank Access Num	Locus	Incyte CloneID	GenBank Id (Incyte)
-119620	prominin (mouse)-like 1	2	NM_006017		2070568	pINCY
-178711	Homo sapiens clone 23809 mRNA sequence	2	AI651122		2947914	pINCY
-178697	dachshund (Drosophila) homolog	2	NM_004392	13q22	2923635	pINCY
-178026	S100 calcium-binding protein P	2	AI151190	4p16	2060823	pSport1
-178434	matrix Gla protein	1.9	AV733652	12p13.1- p12.3	2527665	pINCY
-128154	Microfibril-associated glycoprotein-2	1.9	AI802950	12p13.1- p12.3	148825	pBlue
-123641	ATPase, Ca ⁺⁺ transporting, cardiac muscle, slow twitch 2	1.9	NM_001681	12q23-q24.1	1275854	pINCY
-177424	3-prime-phosphoadenosine 5-prime-phosphosulfate synthase 2	1.9	NM_004670	10q23-q24	1603408	pINCY
-118899	ESTs	1.9	AI133467		1403294	pINCY
-121090	transcobalamin I (vitamin B12 binding protein, R binder family)	1.9	NM_001062	11q11-q12	642823	pSport1
-119591	mitogen-activated protein kinase kinase 7	1.9	NM_005043		3031764	pINCY
-176932	small proline-rich protein 1B (cornifin)	1.9	BE932403	1q21-q22	549728	pSport1
-121551	novel RGD-containing protein	1.9	NM_006571	8p12-p11	2061154	pSport1
-127240	DEK oncogene (DNA binding)	1.8	AI924222	6p23	1980248	pSport1
-177400	snuportin-1	1.9	AI240250		1562658	pINCY

43

GeneID (Incyte)	GeneName	Diff Expr	Genebank Access Num	Locus	Incyte CloneID	GenBank Id (Incyte)
-121281	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)	1.8	AI951454	2p14-p12	603597	pSport1
-121533	UDP-galactose transporter related	1.8	NM_005827		2058073	pSport1
-119109	serologically defined colon cancer antigen 43	1.8	AK000694		1503744	pINCY
-127253	profilin 2	1.8	AI453769	3q25.1-q25.2	1979601	pSport1
-118939	guanylate cyclase 1, soluble, beta 3	1.9	NM_000857	4q31.3-q33	1417408	pINCY
-121315	multimerin	1.9	NM_007351	4q22	781087	pSport1

Table 2

Pterygium vs Conjunctiva (MAC): Category "DECREASE" / Down-regulated

Probe Set ID (Genbank ID)	SLR	Title	Gene Symbol	Map Location	GO bio process	GO cell component	GO molec function	GenMAPP Pathway
201041_s_at (NM_004417.2)	-1.82	dual specificity phosphatase 1	DUSP1	5q34	GO:6979;oxidative stress response; predicted/computed		GO:4726;non-membrane spanning protein tyrosine phosphatase;experimental evidence	
201044_x_at (AA530892)	-4.85	dual specificity phosphatase 1	DUSP1	5q34	GO:6979;oxidative stress response; predicted/computed		GO:4726;non-membrane spanning protein tyrosine phosphatase;experimental evidence	
201466_s_at (NM_002228.2)	-1.46	v-jun sarcoma virus 17 oncogene homolog (avian)	JUN	1p32-p31		GO:228;nuclear chromosome; experimental evidence	GO:3700;transcription factor;predicted/computed GO:3702;RNA polymerase II transcription factor; experimental evidence GO:3677;DNA binding; experimental evidence	Apoptosis MAPK Cascade TGF Beta Signaling Pathway Wnt Signaling Pathway
201473_at (NM_002229.1)	-1.85	jun B proto- oncogene	JUNB	19p13.2	GO:6357;transcription regulation from Pol II promoter;experimental evidence	GO:5717;chromatin;experimental evidence	GO:3702;RNA polymerase II transcription factor; experimental evidence GO:3714;transcription co- repressor;experimental evidence GO:3713;transcription co- activator;experimental evidence GO:3677;DNA binding; experimental evidence	
201531_at (NM_003407.1)	-1.35	zinc finger protein 36, C3H type, homolog (mouse)	ZFP36	19q13.1	GO:6402;mRNA catabolism; experimental evidence	GO:5737;cytoplasm;experimental evidence	GO:3727;single-stranded RNA binding;predicted/computed	
201694_s_at (NM_001964.1)	-2.01	early growth response 1	EGRI	5q31.1			GO:3700;transcription factor;predicted/computed GO:3677;DNA binding; predicted/computed	
202081_at (NM_004907.1)	-1.19	immediate early protein	ETR101	19p13.13				

202340_x_at (NM_002135.1)	-2.68	nuclear receptor subfamily 4, group A, member 1	NR4A1	12q13	GO:7165;signal transduction; predicted/computed		GO:3677;DNA binding; predicted/computed GO:4879;ligand-dependent nuclear receptor; predicted/computed	<u>Nuclear Receptors</u>
202672_s_at (NM_001674.1)	-2.57	activating transcription factor 3	ATF3	1q32.2			GO:3700;transcription factor; experimental evidence GO:3714;transcription co- repressor;experimental evidence	<u>G Protein Signaling Pathways</u>
202768_at (NM_006732.1)	-4.95	FBJ murine osteosarcoma viral oncogene homolog B	FOSB	19q13.32	GO:122;repression of transcription from Pol II promoter;experimental evidence GO:74;cell cycle control;not recorded GO:7275;developmental processes; experimental evidence GO:7610;behavior;experimental evidence		GO:8134;transcription factor binding;predicted/computed GO:3677;DNA binding; experimental evidence	
204285_s_at (AI857639)	-1.83	phorbol-12- myristate-13- acetate-induced protein 1	PMAIP1	18q21.31	GO:7048;oncogenesis;predicted/ computed			
204621_s_at (AI935096)	-2.07	nuclear receptor subfamily 4, group A, member 2	NR4A2	2q22-q23	GO:6960;antimicrobial humoral response;predicted/computed GO:7165;signal transduction;not recorded	GO:5634;nucleus;predicted/ computed	GO:3710;transcription activating factor; predicted/computed GO:4879;ligand-dependent nuclear receptor;predicted/ computed	<u>Nuclear Receptors</u>
204622_x_at (NM_006186.1)	-1.79	nuclear receptor subfamily 4, group A, member 2	NR4A2	2q22-q23	GO:6960;antimicrobial humoral response;predicted/computed GO:7165;signal transduction;not recorded	GO:5634;nucleus;predicted/ computed	GO:3710;transcription activating factor; predicted/computed GO:4879;ligand-dependent nuclear receptor;predicted/ computed	<u>Nuclear Receptors</u>
205910_s_at (NM_001807.1)	-3.30	carboxyl ester lipase (bile salt-stimulated lipase)	CEL	9q34.3	GO:6629;lipid metabolism;not recorded GO:6707;cholesterol catabolism;not recorded			<u>KEGG Bile acid biosynthesis</u> <u>KEGG Glycerolipid metabolism</u>
206393_at (NM_003282.1)	-1.62	troponin I, skeletal, fast	TNNI2	11p15.5	GO:6937;muscle contraction regulation;not recorded			<u>Striated Muscle Contraction</u>

209189_at (BC004490.1)	-4.55	v-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	14q24.3	GO:6306;DNA methylation; predicted/computed GO:6357;transcription regulation from Pol II promoter;predicted/computed GO:7048;oncogenesis;predicted/ computed GO:6366;transcription from Pol II promoter;not recorded GO:6954;inflammatory response; predicted/computed GO:8151;cell growth and maintenance; not recorded	GO:5634;nucleus;experimental evidence	GO:3704;specific RNA polymerase II transcription factor;experimental evidence	
211003_x_at (BC003551.1)	-2.92	transglutaminase 2 (C polypeptide, protein-glutamine- gamma- glutamy/transferase)	TGM2	20q12	GO:6464;protein modification; predicted/computed			
212143_s_at (BF340228)	-1.65	insulin-like growth factor binding protein 3	IGFBP3	7p13-p12	GO:7165;signal transduction; predicted/computed	GO:5615;extracellular space;not recorded	GO:5067;insulin-like growth factor receptor binding protein; experimental evidence	
212225_at (AL516854)	-1.98	putative translation initiation factor	SUI1	17	GO:8151;cell growth and maintenance; experimental evidence GO:6446;translational regulation, initiation;experimental evidence	GO:5737;cytoplasm;predicted/ computed	GO:8135;translation factor; experimental evidence GO:3743;translation initiation factor;predicted/computed GO:8181;tumor suppressor; experimental evidence	Translation Factors
213674_x_at (AI858004)	-2.19	immunoglobulin heavy constant gamma 3 (G3m marker)	IGHG3	14q32.33				
213831_at (X00452.1)	-2.83	major histocompatibility complex, class II, DQ alpha 1	HLA- DQA1	6p21.3	GO:9405;pathogenesis;experimental evidence		GO:3821;class II major histocompatibility complex antigen;experimental evidence	

Table 3

Pterygium vs Conjunctiva (MAC): Category "INCREASE" / Up-regulated

Probe Set ID (Genbank ID)	SLR	Title	Gene Symbol	Map Location	GO bio process	GO cell component	GO molec function	GenMAP Pathway
201438_at (NM_004369.1)	1.62	collagen, type VI, alpha 3	COL6A3	2q37	GO:7517;muscle development; predicted/computed	GO:5589;collagen type VI; experimental evidence	GO:5202;collagen; experimental evidence	Collagen
201852_x_at (AI813758)	1.70	collagen, type III, alpha 1 (Ehlers- Danlos syndrome type IV, autosomal dominant)	COL3A1	2q31	GO:7397;histogenesis and organogenesis;not recorded GO:8015;circulation;predicted/ computed	GO:5586;collagen type III; predicted/computed	collagen;predicted/computed	Collagen Inflammatory Response Pathway
201884_at (NM_004363.1)	1.64	carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	19q13.1-q13.2		GO:5887;integral plasma membrane protein; predicted/computed	GO:8222;tumor antigen; experimental evidence	
202112_at (NM_000552.2)	1.33	von Willebrand factor	VWF	12p13.3	GO:7596;blood coagulation; experimental evidence GO:7599; hemostasis;not recorded GO:7155; cell adhesion;experimental evidence	GO:5803;secretory vesicle; experimental evidence GO:5615;extracellular space;not recorded	GO:5488;ligand binding or carrier;not recorded GO:5194; cell adhesion;experimental evidence	Blood Clotting Cascade
202291_s_at (NM_000900.1)	1.51	matrix Gla protein	MGP	12p13.1-p12.3	GO:1502;cartilage condensation; predicted/computed GO:7048; oncogenesis;predicted/computed GO:1503;ossification;predicted/ computed	GO:5578;extracellular matrix; predicted/computed	GO:5201;extracellular matrix structural protein; predicted/computed GO:8147;structural protein of bone;predicted/computed GO:5509;calcium binding;not recorded	
202310_s_at (K01228.1)	2.23	collagen, type I, alpha 1	COL1A1	17q21.3-q22.1	GO:8544;epidermal differentiation; experimental evidence GO:1501; skeletal development;experimental evidence	GO:5584;collagen type I;not recorded	GO:5202;collagen;not recorded GO:8147;structural protein of bone;not recorded	Collagen Inflammatory Response Pathway
202404_s_at (NM_000089.1)	2.24	collagen, type I, alpha 2	COL1A2	7q22.1	GO:1501;skeletal development; predicted/computed	GO:5584;collagen type I;not recorded	GO:5202;collagen;not recorded GO:8147;structural protein of bone;not recorded	Collagen Inflammatory Response Pathway
205694_at (NM_000550.1)	1.39	tyrosinase-related protein 1	TYRP1	9p23		GO:16021;integral membrane protein;not recorded		ER
208335_s_at (NM_002036.1)	2.27	Duffy blood group	FY	1q21-q22		GO:5886;plasma membrane; experimental evidence GO:5887; integral plasma membrane protein;not recorded	GO:4888;transmembrane receptor;experimental evidence GO:4872;receptor; experimental evidence	"GPCRs, Other" Peptide GPCRs

209047_at (AL518391)	1.78	aquaporin 1 (channel-forming integral protein, 28kDa)	AQP1	7p14	GO:6833;water transport; predicted/computed GO:7588; excretion;predicted/computed	GO:5887;integral plasma membrane protein; predicted/computed	GO:5372;water transporter; predicted/computed	
209771_x_at (AA761181)	1.65	CD24 antigen (small cell lung carcinoma cluster 4 antigen)	CD24	6q21	GO:6959;humoral defense mechanism;predicted/computed	GO:5886;plasma membrane;not recorded		
210297_s_at (U22178.1)	1.79	Microsemino- protein, beta-	MSMB	10q11.2		GO:5615;extracellular space; predicted/computed GO:5634; nucleus;experimental evidence		
210495_x_at (AF130095.1)	2.27	fibronectin 1	FN1	2q34	GO:7165;signal transduction;not recorded GO:7155;cell adhesion; not recorded GO:6928;cell motility; not recorded	GO:5625;soluble fraction;not recorded GO:5615;extracellular space;not recorded GO:5578;extracellular matrix;predicted/computed	GO:5194;cell adhesion;not recorded	
211161_s_at (AF130082.1)	1.84	Clone of FLC 1492Pro 3121 Collagen type 3 alpha 1						
211719_x_at (BC005858.1)	3.04	fibronectin 1	FN1	2q34	GO:7165;signal transduction;not recorded GO:7155;cell adhesion; not recorded GO:6928;cell motility; not recorded	GO:5625;soluble fraction;not recorded GO:5615;extracellular space;not recorded GO:5578; extracellular matrix; predicted/computed	GO:5194;cell adhesion;not recorded	Inflammatory Response Pathway
211964_at (X05610.1)	1.85	collagen, type IV, alpha 2	COL4A2	13q34		GO:5587;collagen type IV;not recorded	GO:5202;collagen;not recorded	Collagen
212190_at (AL541302)	1.31	serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2	SERPINE2	2q33-q35	GO:7275;developmental processes; predicted/computed		GO:4868;serpin;not recorded	
212865_s_at (BF449063)	1.35	collagen, type XIV, alpha 1 (undulin)	COL14A1	8q23				
212950_at (BF941499)	1.38	KIAA0758 protein	KIAA0758	18				"GPCRs, Other"
213905_x_at (AA845258)	1.44	biglycan	BGN	Xq28				
214303_x_at (AW192795)	1.88	mucin 5, subtype B, tracheobronchial	MUC5B	11p15				

215076_s_at (AU144167)	1.86	collagen, type III, alpha 1 (Ehlers- Danlos syndrome type IV, autosomal dominant)	COL3A1	2q31	GO:7397;histogenesis and organogenesis;not recorded GO:8015;circulation;predicted/com puted	GO:5586;collagen type III; predicted/computed	GO:5202;collagen;predicted/ computed	Collagen Inflammatory Response Pathway
216379_x_at (AK000168.1)	1.61	KIAA1919 protein	KIAA1919	6q22.1				
217272_s_at (AJ001698.1)	1.29	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 13	SERPINF 13	18q21.3-q22	GO:9411;UV response; experimental evidence		GO:4868;serpin;predicted/ computed	
218353_at (NM_05226.1)	1.65	MSTP032 protein	MSTP032	1q23.1				
219134_at (NM_022159.1)	4.20	EGF-TM7- latrophilin-related protein	ETL	1p33-p32				"GPCRs, Class B Secretin-like"
221529_s_at (AF326591.1)	1.56	plasmalemma vesicle associated protein	PLVAP	19p13.2				
32128_at (Y13710)	2.43	chemokine (C-C motif) ligand 18 (pulmonary and activation- regulated)	CCL18	17q11.2	GO:6960;antimicrobial humoral response;predicted/computed GO:7165;signal transduction; experimental evidence GO:6955; immune response; predicted/computed GO:6935;chemotaxis;experimental evidence GO:7154;cell communication;predicted/computed GO:7267;cell-cell signalling; experimental evidence GO:9607;response to biotic stimulus;predicted/computed		GO:8009;chemokine; experimental evidence	

CLAIMS

1. A method of monitoring the progression of, or diagnosing, a disease or condition associated with pterygium, said method comprising:

making a quantitative determination of the transcript level of at least one gene shown in Table 1, 2 or 3 in a sample of cells obtained from the site of said disease or condition; and

comparing the transcript level so determined with the transcript level of said at least one gene obtained from a control sample of cells.

2. A method according to claim 1 wherein the control sample is conjunctival cells of the patient whose progression or diagnosis of pterygium is being monitored.

3. A method according to claim 1 wherein the transcript level of a plurality of genes is quantitatively determined.

4. A method according to claim 3 wherein the transcript level of at least 5 genes is quantitatively determined.

5. A method according to claim 3 wherein the transcript level of at least 10 genes is quantitatively determined.

6. A method according to claim 3 wherein the transcript level of at least 20 genes is quantitatively determined.

7. A method according to claim 1 wherein the genes whose transcript levels are determined are selected from the group of genes responsible for cell growth and maintenance, cell communication, cell adhesion, cyclins, translation initiation factors, signal transduction, oncogenes, extra-cellular

matrix, UV-sensitive and cancer-related genes.

8. A method according to claim 7 wherein the genes whose transcript levels are determined are selected from at least two of the categories defined in claim 7.

9. A method according to claim 7 wherein the genes whose transcript levels are determined are selected from at least three of the categories defined in claim 7.

10. A method according to claim 7 wherein the genes whose transcript levels are determined are selected from at least four of the categories defined in claim 7.

11. A method according to claim 7 wherein the genes whose transcript levels are determined are selected from at least five of the categories defined in claim 7.

12. A method according to claim 1 wherein the determination of the transcript level of the at least one gene is by hybridisation to a gene chip array.

13. A method according to claim 12 wherein the gene chip array is from Affymetrix.

14. A gene chip array suitable for use in the method of any one of claims 1 to 13 comprising at least one nucleic acid suitable for detection of at least one gene shown in Table 1, 2 or 3, optionally a control specific for said at least one gene and optionally at least one control for said gene chip.

15. A gene chip array according to claim 14 having from 40 to 8000 nucleic acids.

16. A gene chip array according to claim 14 having from 50 to 400 nucleic acids.

17. A gene chip array according to claim 14 having from 50 to 100 nucleic acids.

18. A gene chip array according to claim 14 comprising nucleic acids specific for detection of at least 10 of the transcripts of Table 1, 2 or 3.

19. A gene chip array according to claim 14 comprising nucleic acids specific for detection of at least 20 of the transcripts of Table 1, 2 or 3.

20. A gene chip array according to claim 14 comprising nucleic acids specific for detection of at least 40 of the transcripts of Table 1, 2 or 3.

21. An assay method for a modulator of pterygium, wherein said method comprises:

- (a) providing a protein encoded by a gene selected from Table 1, 2 or 3;
- (b) bringing said protein into contact with a candidate modulator of its activity; and
- (c) determining whether said candidate modulator is capable of modulating the activity of said protein.

22. An assay method for a modulator of pterygium, wherein said method comprises:

- (a) providing a pterygial cell in culture;
- (b) bringing said cell into contact with a candidate modulator of pterygium; and
- (c) determining whether said candidate modulator is capable

of modulating the transcript level of at least one gene selected from Table 1, 2 or 3.

23. An assay method for a regulator of DNA function comprising providing a gene transcript sequence listed in Table 1, 2 or 3 attached to its native promoter region, bringing this into contact with a candidate modulator under conditions where in the absence of modulator gene the transcript is produced, and determining whether the candidate modulator alters transcription.

24. A modulator obtained by the method of claim 23 for use in a method of modulating pterygial tissue growth in a human patient.

25. A vector comprising an EST sequence from Table 1 linked to a promoter for transcription of said sequence.

26. A vector according to claim 25 for use in therapy.

27. An isolated nucleic acid encoding or identified by an EST of Table 1 or a fragment thereof.

28. An amino acid sequence encoded by the EST sequence of claim 27, or a fragment of such a sequence having at least 15 amino acids, or a peptide having at least 50% sequence identity to said amino acid sequence or fragment thereof.

29. The use of an EST as identified in Table 1 or an amino acid sequence, fragment or peptide sequence as defined in claim 28 in the identification of nucleic acid sequences suitable for use as a marker of pterygium.

30. The use of an amino acid sequence, fragment thereof or peptide as defined in claim 28 in the formulation of a composition for use in prophylactic or therapeutic treatment of pterygium.

31. An antibody capable of binding selectively to an amino acid sequence, fragment or peptide as defined in claim 28.

32. The antibody of claim 31 which is monoclonal.