



(51) International Patent Classification:

C12Q 1/68 (2018.01) *C12Q 1/6827* (2018.01)
C12Q 1/6844 (2018.01) *G01N 33/50* (2006.01)
C12Q 1/6869 (2018.01) *A61P 27/02* (2006.01)

(21) International Application Number:

PCT/US2018/029836

(22) International Filing Date:

27 April 2018 (27.04.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/492,060 28 April 2017 (28.04.2017) US
 62/570,620 10 October 2017 (10.10.2017) US

(71) Applicant: AVELLINO LAB USA, INC. [US/US]; 1505
 Adams Drive, Suite B2, Menlo Park, California 94025 (US).

(72) Inventors: DEDIONISIO, Larry; 1505 Adams Drive,
 Suite B2, Menlo Park, California 94025 (US). MOORE,
 Tara; 1505 Adams Drive, Suite B2, Menlo Park, California
 94025 (US). NESBIT, Andrew; 1505 Adams Drive, Suite
 B2, Menlo Park, California 94025 (US).

(74) Agent: LEE, Janice H. Lee et al.; Morgan, Lewis & Bock-
 ius LLP, 1400 Page Mill Rd., Palo Alto, California 94304
 (US).

(81) Designated States (unless otherwise indicated, for every
 kind of national protection available): AE, AG, AL, AM,
 AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
 CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
 DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
 HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
 KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
 MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
 OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
 SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
 TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
 kind of regional protection available): ARIPO (BW, GH,
 GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
 UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
 TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
 EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
 MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
 TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
 KM, ML, MR, NE, SN, TD, TG).

(54) Title: METHODS FOR DETECTING ALLELES ASSOCIATED WITH KERATOCONUS

Figure 1

Gene	Position	ref	variant	rsid	chr	pos	allele	frequency	frequency
PRPF1	20823478 T	NC_010406.1	G	15172181	7	45222888	15507575	0.01	65.945654
GRN	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
GRN	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
GRN	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	T	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	A	11130623	7	23818912	15507575	0.06	65.945654
PRPF1	20827292 G	NC_010406.1	G	11130623	7	23818912	15507575	0.06	65.945654
PRPF1</									

Published:

— *with international search report (Art. 21(3))*

METHODS FOR DETECTING ALLELES ASSOCIATED WITH KERATOCONUS

FIELD OF THE APPLICATION

[0001] This application generally relates to methods for the isolation and detection of disease-associated genetic alleles. In particular, this application relates to methods for the detection of an alleles associated with keratoconus diagnosis and prognosis.

BACKGROUND

[0002] Keratoconus (KC) is the most common corneal ectatic disorder with approximately 6 – 23.5% of subjects carrying a positive family history (Wheeler, J., Hauser, M.A., Afshari, N.A., Allingham, R.R., Liu, Y., *Reproductive Sys Sexual Disord* 2012; S:6). The reported prevalence of KC ranges from 8.8 to 54.4 per 100,000. This variation in prevalence is partly due to the different criteria used to diagnose the disease. (Wheeler, J., Hauser, M.A., Afshari, N.A., Allingham, R.R., Liu, Y., *Reproductive Sys Sexual Disord* 2012; S:6; and Nowak, D., Gajecka, M., *Middle East Afr J Ophthalmol* 2011; 18(1): 2-6). Many studies exist within the literature that attempt to define the genetic causes of KC. These studies have uncovered numerous possible genetic variants or SNPs that are believed to contribute to the etiology of the disease depending on the experimental parameters.

[0003] KC is a common corneal disorder where the central or paracentral cornea undergoes progressive thinning and steepening causing irregular astigmatism. The hereditary pattern is neither prominent nor predictable, but positive family histories have been reported. The incidence of KC is often reported to be 1 in 2000 people. KC can show the following pathologic findings, including, fragmentation of Bowman's layer, thinning of stroma and overlying epithelium, folds or breaks in Descemet's membrane, and variable amounts of diffuse corneal scarring.

[0004] Histopathology studies demonstrate breaks in or complete absence of Bowman's layer, collagen disorganization, scarring and thinning. The etiology of these changes is not known, though some suspect changes in enzymes that lead to breakdown of collagen in the cornea. While a genetic predisposition to KC is suggested, a specific gene has not been identified. The majority of KC cases are bilateral, but often asymmetric. The less affected eye may show a high amount of astigmatism or mild steepening. Onset is typically in early

adolescence and progresses into the mid-20's and 30's. However, cases may begin much earlier or later in life. There is variable progression for each individual. There is often a history of frequent changes in eye glasses which do not adequately correct vision. Another common progression is from soft contact lenses, to toric or astigmatism correcting contact lens, to rigid gas permeable contact lens.

[0005] No preventive strategy has been proven effective to date. Some feel that eye rubbing or pressure (*e.g.*, sleeping with the hand against the eye) can cause and/or lead to progression of KC, so subjects should be informed not to rub the eyes. In some subjects, avoidance of allergens may help decrease eye irritation and therefore decrease eye rubbing.

[0006] At present, diagnosis can be made by slit-lamp examination and observation of central or inferior corneal thinning. Computerized videokeratography is also useful in detecting early KC and allows following its progression. Ultrasound pachymetry can also be used to measure the thinnest zone on the cornea. New algorithms using computerized videokeratography have been devised which now allow the detection of forme fruste, subclinical or suspected keratoconus. These devices may allow better screening of subjects for prospective refractive surgery, however there remains a need in the art for better prognostic and diagnostic methods.

[0007] The present disclosure meets this need and by providing methods for prognosis and diagnosis of KC by detection of mutated alleles associated with keratoconus.

SUMMARY

[0008] The present disclosure provides improved methods for the detection of one or more alleles associated with KC.

[0009] In some embodiments, the disclosure provides methods for detecting variants related to KC in a subject, the method comprising detecting two or more genetic variants (*e.g.*, single nucleotide polymorphisms (SNPs) and indels) in a sample from a subject, wherein two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 1, and wherein the presence of two or more genetic variants is indicative of KC in the subject.

[0010] In some embodiments, the disclosure provides methods for diagnosing or prognosing KC in a subject, the method comprising detecting two or more genetic variants (*e.g.*, single nucleotide polymorphisms (SNPs) and indels) in a sample from a subject, wherein two or

more genetic variants are selected from the group consisting of genetic variants listed in Figure 1, and wherein the presence of two or more genetic variants is indicative of a diagnosis or prognosis of KC in the subject.

[0011] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 2. In additional embodiments, the subject is Afro-American. In further embodiments, the Afro-American is identified by detecting two or more genetic variants specific to the Afro-American.

[0012] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 3. In additional embodiments, the subject is Caucasian. In further embodiments, the Caucasian is identified by detecting two or more genetic variants specific to the Caucasian.

[0013] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 4. In additional embodiments, the subject is Hispanic. In further embodiments, the Hispanic is identified by detecting two or more genetic variants specific to the Hispanic.

[0014] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 5. In additional embodiments, the subject is East Asian or Korean. In further embodiments, the East Asian or Korean is identified by detecting two or more genetic variants specific to the East Asian or Korean.

[0015] In some embodiments, two or more genetic variants are selected from the group consisting of any combination of the mutations (e.g., genetic variants) described herein (e.g., Figures 1-5).

[0016] In some embodiments, said genetic variant detection is by a sequencing method.

[0017] In some embodiments, the disclosure provides methods for detecting variants related to or causing KC in a subject, the method comprising detecting two or more genetic variants (e.g., single nucleotide polymorphisms (SNPs) and indels) in a sample from a subject, wherein two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 1, and wherein the presence of two or more genetic variants is indicative of KC in the subject.

[0018] In some embodiments, the disclosure provides methods for predicting risk of developing KC in a subject, the method comprising detecting two or more genetic variants in a sample from a subject, wherein the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 1, and wherein the presence of two or more of genetic variants is indicative of risk for the development of KC in the subject.

[0019] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 2. In additional embodiments, the subject is Afro-American.

[0020] In some embodiments, the two or more genetic variants are selected from the group listed in Figure 3. In additional embodiments, the subject is Caucasian.

[0021] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed Figure 4. In additional embodiments, the subject is Hispanic.

[0022] In some embodiments, the two or more genetic variants are selected from the group consisting of listed in Figure 5. In additional embodiments, the subject is East Asian or Korean.

[0023] In some embodiments, the two or more genetic variants are selected from the group consisting of any combination of the mutations (e.g., genetic variants) described herein (e.g., Figures 1-5).

[0024] In some embodiments, said variant detection is by a sequencing method.

[0025] In some embodiments, the disclosure provides methods for developing a treatment regimen for the treatment of KC in a subject, the method comprising detecting two or more genetic variants in a sample from a subject, wherein the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 1, and wherein the presence of two or more genetic variants is indicative of the need for a KC treatment regimen in the subject.

[0026] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 2. In additional embodiments, the subject is Afro-American.

[0027] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 3. In additional embodiments, the subject is Caucasian.

[0028] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 4. In additional embodiments, the subject is Hispanic.

[0029] In some embodiments, the two or more genetic variants are selected from the group consisting of genetic variants listed in Figure 5. In additional embodiments, the subject is East Asian or Korean.

[0030] In some embodiments, the two or more genetic variants (e.g., SNPs) are selected from the group consisting of any combination of the mutations (e.g., genetic variants) described herein (e.g., Figures 1-5).

[0031] In some embodiments, said variant detection is by a sequencing method.

[0032] In some embodiments, the disclosure provides methods for treating keratoconus in a subject, the method comprising diagnosing or prognosing KC and treating KC in the subject. In further embodiments, the treatment may comprise wearing eye glasses or contact lenses, and/or performing collagen cross-linking or corneal transplant.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] **Figure 1** depicts a table listing the frequency of each variant found within the study cohort ordered by chromosome, gene symbol, dbSNP id and ethnicity. The list is divided into shared variants between ethnic groups, Caucasian (C), East Asian (EA), Hispanic (H), African American (AA), and South Asian (SA), followed by variants that are specific to each group. A total of 1,117 nonsynonymous single nucleotide variants (SNVs) and insertion/deletions (INDELs) within 259 genes spanning the entire exome are listed. A RefSeq (ncbi.nlm.nih.gov/) accession number along with the minor allele frequency (MAF) taken from the Exome Aggregation Consortium (ExAC, exac.broadinstitute.org/) is provided. N = total alleles for each group.

[0034] **Figure 2** lists genetic variants specific to Afro-American subjects having keratoconus.

[0035] **Figure 3** lists genetic variants specific to Caucasian subjects having keratoconus.

[0036] **Figure 4** lists genetic variants specific to Hispanic subjects having keratoconus.

[0037] **Figure 5** lists genetic variants specific to East Asian subjects having keratoconus.

[0038] **Figure 6** lists additional genetic variants shared to all subjects having keratoconus.

[0039] **Figure 7** depicts a table that lists an odds ratio (OR) and risk score assignment for rare variants from cornea genes identified within the Caucasian group. Variants were taken from 48 genes related to corneal structure and function and were drawn from a larger list of variants in the Caucasian study cohort. Variants were further selected based on their presence in 1 or more case samples and in 0 ethnic-matched controls. Risk scores were derived from an algorithm incorporating adjusted ORs from conservation priors in a Bayesian model, and also *in silico* predictions from 7 bioinformatic tools indicated by red and yellow circles.

[0040] Abbreviations: A = Afro-American, C = Caucasian, H = Hispanic, and EA = East Asian.

DETAILED DESCRIPTION

[0041] The detection of disease-related variants is an increasingly important tool for the diagnosis and prognosis of various medical conditions. With regard to KC, the present disclosure provides methods for detection of mutant alleles and use of this information in or to diagnose a subject with KC as well as to predict the risk of an individual in developing KC.

[0042] The term “invention” or “present invention” as used herein is not meant to be limiting to any one specific embodiment of the invention but applies generally to any and all embodiments of the invention as described in the claims and specification.

[0043] As used herein, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise. Thus, for example, references to “the method” includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure. It should be understood that the use of “and/or” is defined inclusively such that the term “a, b and/or c” should be read to include the sets of “a,” “b,” “c,” “a and b,” “b and c,” “c and a,” and “a, b and c.”

[0044] As used herein, the term “about” means modifying, for example, lengths of nucleotide sequences, degrees of errors, dimensions, the quantity of an ingredient in a composition, concentrations, volumes, process temperature, process time, yields, flow rates, pressures, and like values, and ranges thereof, refers to variation in the numerical quantity that may occur, for example, through typical measuring and handling procedures used for making compounds, compositions, concentrates or use formulations; through inadvertent error in these procedures; through differences in the manufacture, source, or purity of starting materials or ingredients used to carry out the methods; and like considerations. The term “about” also encompasses amounts that differ due to aging of, for example, a composition, formulation, or cell culture with a particular initial concentration or mixture, and amounts that differ due to mixing or processing a composition or formulation with a particular initial concentration or mixture. Whether modified by the term “about” the claims appended hereto include equivalents to these quantities. The term “about” further may refer to a range of values that are similar to the stated reference value. In certain embodiments, the term “about” refers to a range of values that fall within 50, 25, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 percent or less of the stated reference value.

[0045] As used herein, the term “polymorphism” and variants thereof refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. The terms “genetic mutation” or “genetic variation” and variants thereof include polymorphisms.

[0046] As used herein the term “single nucleotide polymorphism” (“SNP”) and variants thereof refers to a site of one nucleotide that varies between alleles. A single nucleotide polymorphism (SNP) is a single base change or point mutation but variants also include the so-called “indel” mutations (insertions or deletions of 1 to several up to 75 nucleotides), resulting in genetic variation between individuals. SNPs, which make up about 90% of all human genetic variation, occur every 100 to 300 bases along the 3-billion-base human genome. However, SNPs can occur much more frequently in other organisms like viruses. SNPs can occur in coding or non-coding regions of the genome. A SNP in the coding region may or may not change the amino acid sequence of a protein product. A SNP in a non-coding region can alter promoters or processing sites and may affect gene transcription and/or processing. Knowledge of whether an individual has particular SNPs in a genomic region of interest may provide sufficient information to develop diagnostic, preventive and therapeutic applications for a variety of diseases.

[0047] The term “primer” and variants thereof refers to an oligonucleotide that acts as a point of initiation of DNA synthesis in a polymerase chain reaction (PCR). A primer is usually about 10 to about 35 nucleotides in length and hybridizes to a region complementary to the target sequence.

[0048] The term “probe” and variants thereof (*e.g.*, detection probe) refers to an oligonucleotide that hybridizes to a target nucleic acid in a PCR reaction. Target sequence refers to a region of nucleic acid that is to be analyzed and comprises the polymorphic site of interest.

[0049] The hybridization occurs in such a manner that the probes within a probe set may be modified to form a new, larger molecular entity (*e.g.*, a probe product). The probes herein may hybridize to the nucleic acid regions of interest under stringent conditions. As used herein the term “stringency” is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. “Stringency” typically occurs in a range from about T_m ° C to about 20° C to 25° C below T_m . A stringent hybridization may be used to isolate and detect identical polynucleotide sequences or to isolate and detect similar or related polynucleotide sequences. Under “stringent conditions” the nucleotide sequence, in its entirety or portions thereof, will hybridize to its exact complement and closely related sequences. Low stringency conditions comprise conditions equivalent to binding or hybridization at 68° C. in a solution consisting of 5×SSPE (43.8 g/l NaCl, 6.9 g/l $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5× Denhardt’s reagent (50× Denhardt’s contains per 500 ml: 5 g Ficoll (Type 400), 5 g BSA) and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 2.0×SSPE, 0.1% SDS at room temperature when a probe of about 100 to about 1000 nucleotides in length is employed. It is well known in the art that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol), as well as components of the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, conditions which promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the

hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) are well known in the art. High stringency conditions, when used in reference to nucleic acid hybridization, comprise conditions equivalent to binding or hybridization at 68° C in a solution consisting of 5×SSPE, 1% SDS, 5× Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1×SSPE and 0.1% SDS at 68° C when a probe of about 100 to about 1000 nucleotides in length is employed.

[0050] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention pertains. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, various embodiments of methods and materials are specifically described herein.

[0051] As explained above, KC is the most common corneal ectatic disorder with approximately 6 – 23.5% of patients carrying a positive family history. (Wheeler, J., Hauser, M.A., Afshari, N.A., Allingham, R.R., Liu, Y. *Reproductive Sys Sexual Disord* 2012; S:6.) The reported prevalence of KC ranges from 8.8 to 54.4 per 100,000. This variation in prevalence is partly due to the different criteria used to diagnose the disease. (Wheeler, J., Hauser, M.A., Afshari, N.A., Allingham, R.R., Liu, Y. *Reproductive Sys Sexual Disord* 2012; S:6; and Nowak, D., Gajecka, M. *Middle East Afr J Ophthalmol* 2011; 18(1):2-6) Many studies exist within the literature that attempt to define the genetic causes of KC. These studies have uncovered numerous possible genetic variants or SNPs that are believed to contribute to the etiology of the disease depending on the experimental parameters.

[0052] In general, the work conducted thus far primarily makes use of micro-satellite genotyping and micro-chip technologies (SNP arrays) to interrogate regions of interest within the genome. In comparison, the study described herein utilized Next Gen Sequencing (NGS) technology to identify and to validate genetic variants that contribute to the etiology of the disease. The study involved a whole exome sequencing (WES) approach (ACE Platform™; Personalis Inc., Menlo Park, CA) in which the ~22,000 genes that comprise the human exome were captured and sequenced; single point mutations or variants including INDELS were identified.

[0053] It is recognized that within the human genome there exist various loci harboring gene mutations that contribute to the phenotypical profile of KC. Among those loci documented in the literature are regions mapped to chromosomes 15q2.32 and 15q22.33-q24.2, 13q32,

16q22.3-q23.1, 3p14-q13, 5q14.3-q21.1, 5q21.2 and 5q32-q33, 1p36.23-36.21 and 8q13.1-q21.11, 9q34, 14q11.2 and 14q24.3 (see, for example, Bisceglia L, De Bonis P, Pizzicoli C et al., *Invest Ophthalmol Vis Sci* 2009; 50: 1081–1086; Hughes AE, Dash DP, Jackson AJ, Frazer DG, Silvestri G, *Invest Ophthalmol Vis Sci* 2003; 44:5063–5066; Gajicka M, Radhakrishna U, Winters D et al., *Invest Ophthalmol Vis Sci* 2009; 50:1531–1539; Czugala, M., Karolak, J.A., Nowak, D.A., et.al., *European Journal of Human Genetics* 2012; 20:389–397; Tyynismaa H, Sistonen P, Tuupanen S et al.; *Invest Ophthalmol Vis Sci* 2002; 43: 3160–3164; Brancati F, Valente EM, Sarkozy A et al., *J Med Genet* 2004; 41:188–192; Tang YG, Rabinowitz YS, Taylor KD et al., *Genet Med* 2005; 7: 397–405; Burdon KP, Coster DJ, Charlesworth JC et al.; *Hum Genet* 2008; 124:379–386; Li, X., Rabinowitz, Y.S., Tang, Y.G., Picornell, Y., Taylor, K.D., Hu, M., Yang, H.; *Invest Ophthalmol Vis Sci* 2006; 47:3791–3795; and Liskova P, Hysi PG, Waseem N, Ebenezer ND, Bhattacharya SS, Tuft SJ, *Arch Ophthalmol* 2010; 128:1191–1195.)

[0054] As explained above, these studies mostly utilize micro-satellite genotyping in conjunction with array chip technologies to interrogate regions of interest within the genome.

[0055] In addition to the above referenced studies, mutations in the visual system homeobox gene 1 (VSX1) have been identified through the targeted screening of this gene in patients diagnosed with KC. The research conducted on the VSX1 gene so far has not clearly identified a causative agent and in fact, much of the literature presents conflicting results. See, for example, Bisceglia, L., Ciaschetti, M., De Bonis, P., Campo, P.A., Pizzicoli, C., Scala, C., Grifa, M., Ciavarella, P., Delle Noci, N., Vaira, F. et al., *Invest Ophthalmol Vis Sci* 2005; 46:39–45, Heon, E., Greenberg, A., Kopp, K.K., Rootman, D., Vincent, A.L., Billingsley, G., Priston, M., Dorval, K.M., Chow, R.L., McInnes, R.R. et al., *Hum Mol Genet* 2002; 11(9):1029–1036, Tang, Y.G., Picornell, Y., Su, X., Li, X., Yang, H. and Rabinowitz, Y.S. *Cornea* 2008; 27:189–192; Aldave, A.J., Yellore, V.S., Salem, A.K., Yoo, G.L., Rayner, S.A., Yang, H., Tang, G.Y., Piconell, Y., Rabinowitz, Y.S., *Invest Ophthalmol Vis Sci* 2006; 47(7):2820-2; Tanwar, M., Kumar, M., Nayak, B., Pathak, D., Sharma, N., Titiyal, J.S. and Dada, R., *Mol Vis* 2010; 16: 2395–2401; Mok, L.W., Baek, S.J., Joo, C.K., *J Hum Genet* 2008; 53:842–849; Jeoung, J.W., Kim, M.K., Park, S.S., Kim, S.Y., Ko, H.S., Won Ryang Wee, W.R., Jin Hak Lee, J.H., *Cornea* 2012; 31,7:746-750; Dehkordi, F.A., Rashki, A., Bagheri, N., Chaleshtori, M.H., Memarzadeh, E., Salehi, A., Ghatreh, H., Zandi, F., Yazdanpanahi, N., Tabatabaiefar, M.A., Chaleshtori, M.H. *Method. Acta Cytologica* 2013; 57: 646–651; Saeed-Rad, S., Hashemi, H., Mirafteb, M., Noori-Dalooi, M.R., Chaleshtori, M.H.,

Raoofian, R., Jafari, F., Greene, W., Fakhraie, G., Rezvan, F., Heidari, M., *Mol Vis* 2011; 17:3128-3136; Wang, Y., Jin, T., X. Zhang, X., Wei, W., Cui, Y., Geng, T., Liu, Q., Gao, J., Liu, M., Chen, C., Zhang, C., Zhu, X., *Ophthalmic Genetics* 2013; 34,3: 160–166; Dash, D.P., S George, S., O'Prey, D., Burns, D., Nabili, S., Donnelly, U., Hughes, A.E., Silvestri, G., Jackson, J., Frazer, D., Heon, E., Willoughby, C.E., *Eye*, 2010; 24,6: 1085–1092.

[0056] While much investigative work has been carried out on the possible role of the VSX1 gene in the etiology of KC, this is not the only gene that has been targeted for analysis.

[0057] Most prominent among the genes that have been investigated within the literature are the various genes related to the structure of collagen. Collagens are the major protein components of the human cornea, and there exist several types of collagen genes that code for the various collagen proteins. Of interest here are COL4A3 and COL4A4 (Štabuc-Šilih, M., Ravnik-Glavač, M., Glavač, D., Hawlina, M., Stražišar M., *Mol Vis* 2009; 15:2848-2860; Štabuc-Šilih, M., Stražišar, M., Ravnik Glavač, M., Hawlina, Glavač, D.; *Acta Dermatoven APA* 2010; 19(2):3-10; Vitart, V., Bencic, G., Hayward, C., Herman, J.S., Huffman J., Campbell, S., Bucan, K., Navarro, P., Gunjaca, G., Marin, J., Zgaga, L., Kolcic, I., Polasek, O., Kirin, M., Hastie, N.D., Wilson, J.F., Rudan, I., Campbell, H., Vatauvuk, Z., Fleck, B., Wright, A., *Hum Mol Genet* 2010; 19(21): 4304–4311) mapped to 2q36.3 (Vitart, V., Bencic, G., Hayward, C., Herman, J.S., Huffman J., Campbell, S., Bucan, K., Navarro, P., Gunjaca, G., Marin, J., Zgaga, L., Kolcic, I., Polasek, O., Kirin, M., Hastie, N.D., Wilson, J.F., Rudan, I., Campbell, H., Vatauvuk, Z., Fleck, B., Wright, A., *Hum Mol Genet* 2010; 19(21): 4304–4311) along with COL4A1 and COL4A2 mapped to the 13q32 locus (Gajecka M, Radhakrishna U, Winters D et al., *Invest Ophthalmol Vis Sci* 2009; 50:1531–1539; Czugala, M., Karolak, J.A., Nowak, D.A., et.al., *European Journal of Human Genetics* 2012; 20:389–397; Karolak, J.A., Kulinska, K., Nowak, D.M., Pitarque, J.A., Molinari, A., Rydzanicz, M., Bejjani, B.A., Gajecka, M., *Mol Vis* 2011; 17:827-843). In reference to the COL4A3 and COL4A4 genes, Štabuc-Šilih et al. in a study published in 2009 identified several SNPs that carried significant p-values. In this study which included 104 unrelated diagnosed patients and 157 healthy blood donors, polymorphism M1327V located at allele 3979 in the COL4A4 gene had a p-value <0.0001 with 134 point mutations out of 208 total alleles for the cases and 132 out of 314 alleles for the controls (Štabuc-Šilih, M., Ravnik-Glavač, M., Glavač, D., Hawlina, M., Stražišar M., *Mol Vis* 2009; 15:2848-2860). With that said, in a subsequent paper published in 2010, Štabuc-Šilih et al. excludes COL4A3 and COL4A4 from playing a

significant role in KC pathogenesis (Štabuc-Šilih, M., Stražišar, M., Ravnik Glavač, M., Hawlina, Glavač, D., *Acta Dermatoven APA* 2010; 19(2):3-10).

[0058] Similarly, Karolak et al. documents findings relating to the COL4A1 and the COL4A2 genes within Ecuadorian families; 23 individuals from one family, 25 affected individuals from other Ecuadorian families, and 64 Ecuadorian control subjects were included in this study (Karolak, J.A., Kulinska, K., Nowak, D.M., Pitarque, J.A., Molinari, A., Rydzanicz, M., Bejjani, B.A., Gajecka, M., *Mol Vis* 2011; 17:827-843). This study identifies several mutations within the COL4A1 and the COL4A2 genes that were significant. For instance, a polymorphism, Gln1334His found at the 4002 allele on COL4A1 gene was observed more frequently in patients than in healthy individuals in the family where twenty-three individuals ($p=0.056$) were examined. However, there was no difference in the c. 4002A>C allele distribution between the analyzed affected individuals from the remaining KC families and the Ecuadorian control subjects ($p=0.17$).

[0059] In conjunction with the work described above (Karolak, J.A., Kulinska, K., Nowak, D.M., Pitarque, J.A., Molinari, A., Rydzanicz, M., Bejjani, B.A., Gajecka, M., *Mol Vis* 2011; 17:827-843), Czugała et al. conducted a study on the same Ecuadorian family group that revealed eight candidate genes other than COL4A1 and COL4A2 (Czugała, M., Karolak, J.A., Nowak, D.A., et al., *European Journal of Human Genetics* 2012; 20: 389–397). These genes are MBNL1, IPO5, FARP1, RNF113B, STK24, DOCK9, ZIC5 and ZIC2. Ninety-two sequence variants were identified within these eight genes. At least four of the ninety-two variants referred to in this study show a statistical correlation to the KC phenotype. These genes and the SNPs associated with them are located at the 13q32 locus, but another important aspect of both this study and the work conducted with the COL4A1 and COL4A2 genes is that the results are derived from the genetic analysis primarily of one extended family in Ecuador (Czugała, M., Karolak, J.A., Nowak, D.A., et al., *European Journal of Human Genetics* 2012; 20:389–397).

[0060] The case studies referenced here were conducted to further elucidate the role of collagen genes and the role they play within the cornea and to investigate the role of the 13q32 locus, a location on the genome that could be an important hotspot within the human genome (Gajecka M, Radhakrishna U, Winters D et al., *Invest Ophthalmol Vis Sci* 2009; 50:1531–1539; and Czugała, M., Karolak, J.A., Nowak, D.A., et al., *European Journal of Human Genetics* 2012; 20:389–397). COL4A3 and COL4A4 genes, which are known to be

deregulated in KC patients, are often subjected to chromosomal aberrations, and could also be responsible for a decrease in collagen types I and III, a feature often detected in the disease (Critchfield, J.W., Calandra, A.J., Nesburn, A.B., Kenney, M.C., *Exp Eye Res* 1988; **46**: 953-63; Kenney, M.C., Nesburn, A.B., Burgeson, R.E., Butkowsky, R.J., Ljubimov A.V., *Cornea* 1997; **16**:345-51; Meek, K.M., Tuft, S.J., Huang, Y., Gill P.S., Hayes, S., Newton, R.H., Bron, A.J., *Invest Ophthalmol Vis Sci* 2005; **46**:1948-56; Bochert, A., Berlau, J., Koczan, D., Seitz, B., Thiessen, H.J., Guthoff, R.F., *Ophthalmologie* 2003; **100**:545-9; Stachs, O., Bocher, A., Gerber, T., Koczan, D., Thiessen, H.J., Guthoff, R.F., *Ophthalmologie* 2004; **101**:384-9; Pettenati, M.J., Sweatt, A.J., Lantz, P., Stanton, C.A., Reynolds, J., Rao, P.N., Davis, R.M., *Hum Genet* 1997; **101**:26-9).

[0061] The search for a genetic link that defines the subset of KC, labeled as familial KC mostly results in the identification of different SNP candidates depending on the family pedigree. For example, the gene, VSX1 was thought to be a primary candidate based on a few isolated family studies (Bisceglia, L., Ciaschetti, M., De Bonis, P., Campo, P.A., Pizzicoli, C., Scala, C., Grifa, M., Ciavarella, P., Delle Noci, N., Vaira, F. et al., *Invest Ophthalmol Vis Sci* 2005; **46**: 39–45; Heon, E., Greenberg, A., Kopp, K.K., Rootman, D., Vincent, A.L., Billingsley, G., Priston, M., Dorval, K.M., Chow, R.L., McInnes, R.R. et al., *Hum Mol Genet* 2002; **11**(9):1029–1036); however, non-family based studies have also been conducted with this gene that involved unrelated individuals of different ethnicities and geographic locations. These studies attempt to identify specific SNPs within the gene that would better define the role of VSX1 (Aldave, A.J., Yellore, V.S., Salem, A.K., Yoo, G.L., Rayner, S.A., Yang, H., Tang, G.Y., Piconell, Y., Rabinowitz, Y.S., *Invest Ophthalmol Vis Sci* 2006; **47**,7:2820-2; Tanwar, M., Kumar, M., Nayak, B., Pathak, D., Sharma, N., Titiyal, J.S. and Dada, R., *Mol Vis* 2010; **16**: 2395–2401; Mok, L.W., Baek, S.J., Joo, C.K., *J Hum Genet* 2008; **53**: 842–849; Jeoung, J.W., Kim, M.K., Park, S.S., Kim, S.Y., Ko, H.S., Won Ryang Wee, W.R., Jin Hak Lee, J.H., *Cornea* 2012; **31**,7: 746-750; Dehkordi, F.A., Rashki, A., Bagheri, N., Chaleshtori, M.H., Memarzadeh, E., Salehi, A., Ghatreh, H., Zandi, F., Yazdanpanahi, N., Tabatabaiefar, M.A., Chaleshtori, M.H., *Acta Cytologica* 2013; **57**: 646–651, Wang, Y., Jin, T., X. Zhang, X., Wei, W., Cui, Y., Geng, T., Liu, Q., Gao, J., Liu, M., Chen, C., Zhang, C., Zhu, X., *Ophthalmic Genetics* 2013; **34**,3: 160–166; Dash, D.P., S George, S., O’Prey, D., Burns, D., Nabili, S., Donnelly, U., Hughes, A.E., Silvestri, G., Jackson, J., Frazer, D., Heon, E., Willoughby, C.E., *Eye*, 2010; **24**,6: 1085–1092). In general, publications resulting from these studies are inconclusive and in fact, the pathogenic role of

certain non-synonymous candidate SNPs found within the VSX1 gene has been refuted (Tang, Y.G., Picornell, Y., Su, X., Li, X., Yang, H. and Rabinowitz, Y.S., *Cornea* 2008; 27: 189–192; Aldave, A.J., Yellore, V.S., Salem, A.K., Yoo, G.L., Rayner, S.A., Yang, H., Tang, G.Y., Piconell, Y., Rabinowitz, Y.S., *Invest Ophthalmol Vis Sci* 2006; 47(7): 2820-2; Tanwar, M., Kumar, M., Nayak, B., Pathak, D., Sharma, N., Titiyal, J.S. and Dada, R. VSX1 gene analysis in keratoconus. *Mol Vis* 2010; 16: 2395–2401).

[0062] KC with no family associations is the most common form of the disease seen by practicing clinicians (Rabinowitz, Y.S., *Ophthalmol Clin N Am.* 2003; 16(4): 607–620). With that said, it is likely that familial aggregation has been underreported due to undetected forms of KC. Recent advances in diagnostic techniques such as videokeratography may help better understand whether other forms of the disease are, in actuality, inherited.

[0063] The work described above that involves the VSX1 gene (Bisceglia, L., Ciaschetti, M., De Bonis, P., Campo, P.A., Pizzicoli, C., Scala, C., Grifa, M., Ciavarella, P., Delle Noci, N., Vaira, F. et al., *Invest Ophthalmol Vis Sci* 2005; 46:39–45; Heon, E., Greenberg, A., Kopp, K.K., Rootman, D., Vincent, A.L., Billingsley, G., Priston, M., Dorval, K.M., Chow, R.L., McInnes, R.R. et al.; *Hum Mol Genet* 2002; 11(9):1029–1036; Tang, Y.G., Picornell, Y., Su, X., Li, X., Yang, H. and Rabinowitz, Y.S. *Cornea* 2008; 27:189–192; Aldave, A.J., Yellore, V.S., Salem, A.K., Yoo, G.L., Rayner, S.A., Yang, H., Tang, G.Y., Piconell, Y., Rabinowitz, Y.S. *Invest Ophthalmol Vis Sci* 2006; 47(7): 2820-2; Tanwar, M., Kumar, M., Nayak, B., Pathak, D., Sharma, N., Titiyal, J.S. and Dada, R. *Mol Vis* 2010; 16: 2395–2401; Mok, L.W., Baek, S.J., Joo, C.K. *J Hum Genet* 2008; 53: 842–849; Jeoung, J.W., Kim, M.K., Park, S.S., Kim, S.Y., Ko, H.S., Won Ryang Wee, W.R., Jin Hak Lee, J.H. VSX1 Gene and Keratoconus: Genetic Analysis in Korean Patients *Cornea* 2012; 31(7): 746-750; Dehkordi, F.A., Rashki, A., Bagheri, N., Chaleshtori, M.H., Memarzadeh, E., Salehi, A., Ghatreh, H., Zandi, F., Yazdanpanahi, N., Tabatabaiefar, M.A., Chaleshtori, M.H., *Acta Cytologica* 2013; 57: 646–651; Saeed-Rad, S., Hashemi, H., Miraftab, M., Noori-Daloii, M.R., Chaleshtori, M.H., Raoofian, R., Jafari, F., Greene, W., Fakhraie, G., Rezvan, F., Heidari, M. *Mol Vis* 2011; 17: 3128-3136; Wang, Y., Jin, T., X. Zhang, X., Wei, W., Cui, Y., Geng, T., Liu, Q., Gao, J., Liu, M., Chen, C., Zhang, C., Zhu, X., Common single nucleotide polymorphisms and keratoconus in the Han Chinese population. *Ophthalmic Genetics* 2013; 34(3):160–166) and the various COL genes (Štabuc-Šilih, M., Ravnik-Glavač, M., Glavač, D., Hawlina, M., Stražišar M., *Mol Vis* 2009; 15:2848-2860; Štabuc-Šilih, M., Stražišar, M., Ravnik Glavač, M., Hawlina, Glavač, D. *Acta Dermatoven APA* 2010; 19(2):3-10; Karolak, J.A., Kulinska,

K., Nowak, D.M., Pitarque, J.A., Molinari, A., Rydzanicz, M., Bejjani, B.A., Gajecka, M., *Mol Vis* 2011; 17: 827-843; Critchfield, J.W., Calandra, A.J., Nesburn, A.B., Kenney, M.C., *Exp Eye Res* 1988; 46:953-63; Kenney, M.C., Nesburn, A.B., Burgeson, R.E., Butkowsky, R.J., Ljubimov A.V., *Cornea* 1997; 16: 345-51; Meek, K.M., Tuft, S.J., Huang, Y., Gill P.S., Hayes, S., Newton, R.H., Bron, A.J., *Invest Ophthalmol Vis Sci* 2005; 46:1948-56; Bocher, A., Berlau, J., Koczan, D., Seitz, B., Thiessen, H.J., Guthoff, R.F., *Ophthalmologie* 2003; 100:545-9; Stachs, O., Bocher, A., Gerber, T., Koczan, D., Thiessen, H.J., Guthoff, R.F., *Ophthalmologie* 2004; 101: 384-9; Pettenati, M.J., Sweatt, A.J., Lantz, P., Stanton, C.A., Reynolds, J., Rao, P.N., Davis, R.M., *Hum Genet* 1997; 101:26-9; Li, X., Bykhovskaya, Y., Caiado Canedo, A.L., Haritunians, T., Siscovick, D., Anthony J. Aldave, A.J., Szczotka-Flynn, L., Iyengar, S.K., Rotter, J.I., Taylor, K.D., Yaron S. Rabinowitz, Y.S., *Invest Ophthalmol Vis Sci* 2013; 54: 2696–2704) are just a few examples where mutations within genes may be contributing to the phenotype of the disease. These studies primarily focus on the structure and function of one or two genes of interest and in doing so overlook the possibility of other gene mutations within the genome that may contribute to the etiology of the disease. Much of the literature stipulates that genetically, KC is a complex disease (Bisceglia L, De Bonis P, Pizzicoli C et al., *Invest Ophthalmol Vis Sci* 2009; 50:1081–1086; Tang YG, Rabinowitz YS, Taylor KD et al., *Genet Med* 2005; 7:397–405; Li, X., Rabinowitz, Y.S., Tang, Y.G., Picornell, Y., Taylor, K.D., Hu, M., Yang, H., *Invest Ophthalmol Vis Sci* 2006; 47:3791–3795; Liskova P, Hysi PG, Waseem N, Ebenezer ND, *Arch Ophthalmol* 2010; 128:1191–1195; Wheeler, J., Hauser, M.A., Afshari, N.A., Allingham, R.R., Liu, Y., *Reproductive Sys Sexual Disord* 2012; S:6; Nowak, D., Gajecka, M., *Middle East Afr J Ophthalmol* 2011; 18(1):2-6; Burdon, K.P. and Vincent, A.L. *Clin Exp Optom* 2013; 96: 146–154), implicating multiple mutations within more than one gene. HGF and LOX genes harbor SNPs that have been identified as significant in patients diagnosed with KC (Burdon, K.P., Macgregor, S., Bykhovskaya, Y., Javadiyan, S., Li, X., Laurie, K.J., Muszynska, D., Lindsay, R., Lechner, J., Haritunians, T., Henders, A.K., Dash, D., Siscovick, D., Anand, S., Aldave, A., Coster, D.J., Szczotka-Flynn, L., Mills, R.A., Iyengar, S.K., Taylor, K.D., Phillips, T., Grant W. Montgomery, G.W., Rotter, J.I., Hewitt, A.W., Sharma, S., Rabinowitz, Y.S., Willoughby, C., Craig, J.E., *Invest Ophthalmol Vis Sci* 2011; 52(11): 8514-8519; Sahebjada, S., Schache, M., Richardson, A.J., Snibson, G., Daniell, M., Baird, P.N., *PLoS ONE* 2014; 9,1; Dudakova, L., Palos, M., Jirsova, K., Stranecky, V., Krepelova, A., Hysi P.G., Liskova, P., *Eur J Hum Genet*. 2015; Bykhovskaya, Y., Li, X., Epifantseva, I., Haritunians, T., Siscovick, D., Aldave, A., Szczotka-Flynn, L., Iyengar, S.K., Taylor, K.D.,

Rotter, J.I., Rabinowitz, Y.S., *Invest Ophthalmol Vis Sci*; 2012; 53(7): 4152-4157; Hao XD1, Chen P, Chen ZL, Li SX, Wang Y., *Ophthalmic Genet.* 2015; 36(2): 132-136).

[0064] The HGF gene is known to be expressed in the cornea by all three cellular layers (Wilson SE, Walker JW, Chwang EL, He YG., *Invest Ophthalmol Vis Sci.* 1993; 34,8: 2544–2561). The protein is also produced in the lacrimal glands, and HGF expression in corneal keratinocytes is upregulated in response to corneal injury suggesting its involvement in the epithelial wound healing process (Burdon, K.P., Macgregor, S., Bykhovskaya, Y., Javadiyan, S., Li, X., Laurie, K.J., Muszynska, D., Lindsay, R., Lechner, J., Haritunians, T., Henders, A.K., Dash, D., Siscovick, D., Anand, S., Aldave, A., Coster, D.J., Szczotka-Flynn, L., Mills, R.A., Iyengar, S.K., Taylor, K.D., Phillips, T., Grant W. Montgomery, G.W., Rotter, J.I., Hewitt, A.W., Sharma, S., Rabinowitz, Y.S., Willoughby, C., Craig, J.E., *Invest Ophthalmol Vis Sci* 2011; 52(11): 8514-8519; Li Q, Weng J, Mohan RR, et al., *Invest Ophthalmol Vis Sci.* 1996; 37(5): 727–739). Furthermore, certain SNPs associated with the HGF gene have been correlated to hypermetropia and myopia (Yanovitch, T., Li, Y.J., Metlapally, R., Abbott, D., Tran Viet, K.N., Young, T.L., *Mol Vis* 2009; **15**: 1028-1035; Veerappan, S., Pertile, K.K., Islam, A.F., Schäche, M., Chen, C.Y., Mitchell, P., Dirani, M., Baird, P.N., *Ophthalmology* 2010; 117(2): 239–245) along with primary angle closure glaucoma (PACG) (Awadalla, M.S., Thapa, S.S., Burdon, K.P., Hewitt, A.W., Craig, J.E., *Mol Vis* 2011; 17: 2248-2254).

[0065] A subset of the SNPs found to be associated with these various eye conditions were also found in the genomes of KC patients (Burdon, K.P., Macgregor, S., Bykhovskaya, Y., Javadiyan, S., Li, X., Laurie, K.J., Muszynska, D., Lindsay, R., Lechner, J., Haritunians, T., Henders, A.K., Dash, D., Siscovick, D., Anand, S., Aldave, A., Coster, D.J., Szczotka-Flynn, L., Mills, R.A., Iyengar, S.K., Taylor, K.D., Phillips, T., Grant W. Montgomery, G.W., Rotter, J.I., Hewitt, A.W., Sharma, S., Rabinowitz, Y.S., Willoughby, C., Craig, J.E., *Invest Ophthalmol Vis Sci* 2011; 52(11): 8514-8519).

[0066] Regarding the role of the HGF protein in the eye, Burdon et al. states, “The refractive power of the eye is determined at least in part by the shape of the cornea, which is severely altered in KC, thus suggesting overlap between the genetic determinants of these complex ophthalmic conditions” (Burdon, K.P., Macgregor, S., Bykhovskaya, Y., Javadiyan, S., Li, X., Laurie, K.J., Muszynska, D., Lindsay, R., Lechner, J., Haritunians, T., Henders, A.K., Dash, D., Siscovick, D., Anand, S., Aldave, A., Coster, D.J., Szczotka-Flynn, L., Mills, R.A.,

Iyengar, S.K., Taylor, K.D., Phillips, T., Grant W. Montgomery, G.W., Rotter, J.I., Hewitt, A.W., Sharma, S., Rabinowitz, Y.S., Willoughby, C., Craig, J.E., *Invest Ophthalmol Vis Sci* 2011; 52(11): 8514-8519). There exist at least two other studies published within the literature that provide verification that the HGF gene is associated with KC (Sahebjada, S., Schache, M., Richardson, A.J., Snibson, G., Daniell, M., Baird, P.N. *PLoS ONE* 2014, 9(1); Dudakova, L., Palos, M., Jirsova, K., Stranecky, V., Krepelova, A., Hysi P.G., Liskova, P., *Eur J Hum Genet.* 2015).

[0067] LOX encodes an enzyme that initiates the crosslinking of collagens and elastin in a variety of tissues including the cornea (Hamalainen, E.R, Jones, T.A., Sheer, D., Taskinen, K., Pihlajanemi, T., Kivirikko, K.I. *Genomics.* 1991; 11:508–516). Li et al. carried out a genome-wide linkage scan that mapped several loci to KC including the 5q23.2 locus where the LOX gene is located (Li, X., Rabinowitz, Y.S., Tang, Y.G., Picornell, Y., Taylor, K.D., Hu, M., Yang, H. *Invest Ophthalmol Vis Sci* 2006; 47:3791–3795). In addition LOX expression levels were found to be upregulated in a study that analyzed KC epithelium on microarrays (Nielsen, K., Birkenkamp-Demtroder, K., Ehlers, N., Orntoft, T.F. *Invest Ophthalmol Vis Sci.* 2003; 44: 2466–2476). Bykhovskaya et al. in a study that involved two independent panels of patients with KC and controls and KC families found at least four SNPs within this gene that are associated with KC (Bykhovskaya, Y., Li, X., Epifantseva, I., Haritunians, T., Siscovick, D., Aldave, A., Szczotka-Flynn, L., Iyengar, S.K., Taylor, K.D., Rotter, J.I., Rabinowitz, Y.S. *Invest Ophthalmol Vis Sci*; 2012; 53,7: 4152-4157). This work was duplicated in a group that found the rs2956540 SNP to be associated with KC in a population of European descent (Dudakova, L., Palos, M., Jirsova, K., Stranecky, V., Krepelova, A., Hysi P.G., Liskova, P., *Eur J Hum Genet.* 2015) and again in a study conducted on a Han Chinese population (Hao XD1, Chen P, Chen ZL, Li SX, Wang Y., *Ophthalmic Genet.* 2015; 36,2: 132-136).

[0068] Since riboflavin/ultraviolet-a–induced corneal collagen cross-linking (CXL) has become a prevalent form of treatment for the KC patient (Ashwin, P.T., McDonnell, P.J. Collagen cross-linkage: a comprehensive review and directions for future research. *Br J Ophthalmol.* 2010; 94: 965–970), there is interest in a gene such as LOX which encodes for a molecular pathway that lead to collagen cross-linking in the cornea. It is believed that knowing the genotype of the LOX gene within the KC patient may have implications and provide insight into the outcome of CXL treatment (Bykhovskaya, Y., Li, X., Epifantseva, I.,

Haritunians, T., Siscovick, D., Aldave, A., Szczotka-Flynn, L., Iyengar, S.K., Taylor, K.D., Rotter, J.I., Rabinowitz, Y.S. *Invest Ophthalmol Vis Sci*; 2012; 53,7: 4152-4157).

[0069] In one aspect, the disclosure provides methods for isolating genomic samples to identify and validate single nucleotide polymorphism detection. In some embodiments, the genomic samples may be selected from the group consisting of isolated cells, whole blood, serum, plasma, urine, saliva, sweat, fecal matter, and tears.

[0070] In some embodiments, the genomic sample is plasma or serum, and the method further comprises isolating the plasma or serum from a blood sample of the subject.

[0071] In some embodiments, the method includes providing a sample of cells from a subject. In some embodiments, the cells are collected by contacting a cellular surface of a subject with a substrate capable of reversibly immobilizing the cells onto a substrate.

[0072] The disclosed methods are applicable to a variety of cell types obtained from a variety of samples. In some embodiments, the cell type for use with the disclosed methods include but is not limited to epithelial cells, endothelial cells, connective tissue cells, skeletal muscle cells, endocrine cells, cardiac cells, urinary cells, melanocytes, keratinocytes, blood cells, white blood cells, buffy coat, hair cells (including, *e.g.*, hair root cells) and/or salivary cells. In some embodiments, the cells are epithelial cells. In some embodiments, the cells are subcapsular-perivascular (epithelial type 1); pale (epithelial type 2); intermediate (epithelial type 3); dark (epithelial type 4); undifferentiated (epithelial type 5); and large-medullary (epithelial type 6). In some embodiments, the cells are buccal epithelial cells (*e.g.*, epithelial cells collected using a buccal swab). In some embodiments, the sample of cells used in the disclosed methods include any combination of the above identified cell types.

[0073] In some embodiments, the method includes providing a sample of cells from a subject. In some embodiments, the cells provided are buccal epithelial cells.

[0074] The cell sample is collected by any of a variety of methods which allow for reversible binding of the subject's cells to the substrate. In some embodiments, the substrate is employed in a physical interaction with the sample containing the subject's cells in order to reversibly bind the cells to the substrate. In some embodiments, the substrate is employed in a physical interaction with the body of the subject directly in order to reversibly bind the cells to the substrate. In some embodiments, the sample is a buccal cell sample and the sample of buccal cells is collected by contacting a buccal membrane of the subject (*e.g.*, the inside of

their cheek) with a substrate capable of reversibly immobilizing cells that are dislodged from the membrane. In such embodiments, the swab is rubbed against the inside of the subject's cheek with a force equivalent to brushing a person's teeth (*e.g.*, a light amount of force or pressure). Any method which would allow the subject's cells to be reversibly bound to the substrate is contemplated for use with the disclosed methods.

[0075] In some embodiments, the sample is advantageously collected in a non-invasive manner. As such sample collection is accomplished anywhere and by almost anyone. For example, in some embodiments, the sample is collected at a physician's office, at a subject's home, or at a facility where a medical procedure is performed or to be performed. In some embodiments the subject, the subject's doctor, nurses or a physician's assistant or other clinical personnel collects the sample.

[0076] In some embodiments the substrate is made of any of a variety of materials to which cells are reversibly bound. Exemplary substrates include those made of rayon, cotton, silica, an elastomer, a shellac, amber, a natural or synthetic rubber, cellulose, BAKELITE, NYLON, a polystyrene, a polyethylene, a polypropylene, a polyacrylonitrile, or other materials or combinations thereof. In some embodiments, the substrate is a swab having a rayon tip or a cotton tip.

[0077] In some embodiments, the substrate containing the sample is freeze-thawed one or more times (*e.g.*, after being frozen, the substrate containing the sample is thawed, used according to the present methods and re-frozen) and or used in the present methods.

[0078] In another aspect, a variety of lysis solutions have been described and are known to those of skill in the art. Any of these well-known lysis solutions can be employed with the present methods in order to isolate nucleic acids from a sample. Exemplary lysis solutions include those commercially available, such as those sold by INVITROGEN®, QIAGEN®, LIFE TECHNOLOGIES® and other manufacturers, as well as those which can be generated by one of skill in a laboratory setting. Lysis buffers have also been well described and a variety of lysis buffers can find use with the disclosed methods, including for example those described in Molecular Cloning (three volume set, Cold Spring Harbor Laboratory Press, 2012) and Current Protocols (Genetics and Genomics; Molecular Biology; 2003-2013), both of which are incorporated herein by reference for all purposes.

[0079] Cell lysis is a commonly practiced method for the recovery of nucleic acids from within cells. In many cases, the cells are contacted with a lysis solution, commonly an alkaline solution comprising a detergent, or a solution of a lysis enzyme. Such lysis solutions typically contain salts, detergents and buffering agents, as well as other agents that one of skill would understand to use. After full and/or partial lysis, the nucleic acids are recovered from the lysis solution.

[0080] In some embodiments, cells are resuspended in an aqueous buffer, with a pH in the range of from about pH 4 to about 10, about 5 to about 9, about 6 to about 8 or about 7 to about 9.

[0081] In some embodiments, the buffer salt concentration is from about 10 mM to about 200 mM, about 10 mM to about 100 mM or about 20 mM to about 80 mM.

[0082] In some embodiments, the buffer further comprises chelating agents such as ethylenediaminetetraacetic acid (EDTA) or ethylene glycol tetraacetic acid (EGTA).

[0083] In some embodiments, the lysis solution further comprises other compounds to assist with nucleic acid release from cells such as polyols, including for example but not limited to sucrose, as well as sugar alcohols such as maltitol, sorbitol, xylitol, erythritol, and/or isomalt. In some embodiments, polyols are in the range of from about 2% to about 15% w/w, or about 5% to about 15% w/w or about 5% to about 10% w/w.

[0084] In some embodiments, the lysis solutions further comprises surfactants, such as for example but not limited to Triton X-100, SDS, CTAB, X-114, CHAPS, DOC, and/or NP-40. In some embodiments such surfactants are in the range of from about 1% to about 5% w/w, about 1% to about 4% w/w, or about 1% to about 3% w/w.

[0085] In embodiments, the lysis solution further comprises chaotropes, such as for example but not limited to urea, sodium dodecyl sulfate and/or thiourea. In some embodiments, the chaotrope is used at a concentration in the range of from about 0.5 M to 8 M, about 1 M to about 6 M, about 2 M to about 6 M or about 1 M to 3 M.

[0086] In some embodiments, the lysis solution further comprises one or more additional lysis reagents and such lysis reagents are well known in the art. In some embodiments, such lysis reagents include cell wall lytic enzymes, such as for example but not limited to

lysozyme. In some embodiments, lysis reagents comprise alkaline detergent solutions, such as 0.1 aqueous sodium hydroxide containing 0.5% sodium dodecyl sulphate.

[0087] In some embodiments, the lysis solution further comprises aqueous sugar solutions, such as sucrose solution and chelating agents such as EDTA, for example the STET buffer. In certain embodiments, the lysis reagent is prepared by mixing the cell suspension with an equal volume of lysis solution having twice the desired concentration (for example 0.2 sodium hydroxide, 1.0% sodium dodecyl sulphate).

[0088] In some embodiments, after the desired extent of lysis has been achieved, the mixture comprising lysis solution and lysed cells is contacted with a neutralizing or quenching reagent to adjust the conditions such that the lysis reagent does not adversely affect the desired product. In some embodiments, the pH is adjusted to a pH of from about 5 to about 9, about 6 to about 8, about 5 to about 7, about 6 to about 7 or about 6.5 to 7.5 to minimize and/or prevent degradation of the cell contents, including for example but not limited to the nucleic acids. In some embodiments, when the lysis reagent comprises an alkaline solution, the neutralizing reagent comprises an acidic buffer, for example an alkali metal acetate/acetic acid buffer. In some embodiments, lysis conditions, such as temperature and composition of the lysis reagent are chosen such that lysis is substantially completed while minimizing degradation of the desired product, including for example but not limited to nucleic acids.

[0089] Any combination of the above can be employed by one of skill, as well as combined with other known and routine methods, and such combinations are contemplated by the present invention.

[0090] In another aspect, the nucleic acids, including for example but not limited to genomic DNA, are isolated from lysis buffer prior to performing subsequent analysis. In some embodiments, the nucleic acids are isolated from the lysis buffer prior to the performance of additional analyses, such as for example but not limited to real-time PCR analyses. Any of a variety of methods useful in the isolation of small quantities of nucleic acids are used by various embodiments of the disclosed methods. These include but are not limited to precipitation, gel filtration, density gradients and solid phase binding. Such methods have also been described in for example, Molecular Cloning (three volume set, Cold Spring Harbor Laboratory Press, 2012) and Current Protocols (Genetics and Genomics; Molecular Biology; 2003-2013), incorporated herein by reference for all purposes.

[0091] Nucleic Acid precipitation is a well know method for isolation that is known by those of skill in the art. A variety of solid phase binding methods are also known in the art including but not limited to solid phase binding methods that make use of solid phases in the form of beads (e.g., silica, magnetic), columns, membranes or any of a variety other physical forms known in the art. In some embodiments, solid phases used in the disclosed methods reversibly bind nucleic acids. Examples of such solid phases include so-called “mixed-bed” solid phases are mixtures of at least two different solid phases, each of which has a capacity to nucleic acids under different solution conditions, and the ability and/or capacity to release the nucleic acid under different conditions; such as those described in US Patent Application No. 2002/0001812, incorporated by reference herein in its entirety for all purposes. Solid phase affinity for nucleic acids according to the disclosed methods can be through any one of a number of means typically used to bind a solute to a substrate. Examples of such means include but are not limited to, ionic interactions (e.g., anion-exchange chromatography) and hydrophobic interactions (e.g., reversed-phase chromatography), pH differentials and changes, salt differentials and changes (e.g., concentration changes, use of chaotropic salts/agents). Exemplary pH based solid phases include but are not limited to those used in the INVITROGEN ChargeSwitch Normalized Buccal Kit magnetic beads, to which bind nucleic acids at low pH (<6.5) and releases nucleic acids at high pH (>8.5) and mono-amino-N-aminoethyl (MANAE) which binds nucleic acids at a pH of less than 7.5 and release nucleic acids at a pH of greater than 8. Exemplary ion exchange based substrates include but are not limited to DEA-SEPHAROSE™, Q-SEPHAROSE™, and DEAE-SEPHADEX™ from PHARMACIA (Piscataway, N.J.), DOWEX® I from The Dow Chemical Company (Midland, Mich.), AMBERLITE® from Rohm & Haas (Philadelphia, Pa.), DUOLITE® from Duolite International, In. (Cleveland, Ohio), DIALON TI and DIALON TII.

[0092] Any individual method is contemplated for use alone or in combination with other methods, and such useful combination are well known and appreciated by those of skill in the art.

[0093] In another aspect, the disclosed methods are used to isolate nucleic acids, such as genomic DNA (gDNA) for a variety of nucleic acid analyses, including genomic analyses. In some embodiments, such analysis includes detection of variety of genetic mutations, which include but are not limited to deletions, insertions, transitions and transversions. In some embodiments, the mutation is a single-nucleotide polymorphism (SNP).

[0094] A variety of methods for analyzing such isolated nucleic acids, for example but not limited to genomic DNA (gDNA) are known in the art and include nucleic acid sequencing methods (including Next Generation Sequencing methods), PCR methods (including real-time PCR analysis, microarray analysis, hybridization analysis) as well as any other nucleic acid sequence analysis methods that are known in the art, which include a variety of other methods where nucleic acid compositions are analyzed and which are known to those of skill in the art. *See, for example*, Molecular Cloning (three volume set, Cold Spring Harbor Laboratory Press, 2012) and Current Protocols (Genetics and Genomics; Molecular Biology; 2003-2013).

[0095] In one aspect, the SNP described herein may be detected by sequencing. For example, High-throughput or Next Generation Sequencing (NGS) represents an attractive option for detecting mutations within a gene. Distinct from PCR, microarrays, high-resolution melting and mass spectrometry, which all indirectly infer sequence content, NGS directly ascertains the identity of each base and the order in which they fall within a gene. The newest platforms on the market have the capacity to cover an exonic region 10,000 times over, meaning the content of each base position in the sequence is measured thousands of different times. This high level of coverage ensures that the consensus sequence is extremely accurate and enables the detection of rare variants within a heterogeneous sample. For example, in a sample extracted from formalin-fixed, paraffin-embedded (FFPE) tissue, often a mutation of interest is only present at a frequency of 1%. When this sample is sequenced at 10,000X coverage, then even the rare allele, comprising only 1% of the sample, is uniquely measured 100 times over. Thus, NGS provides reliably accurate results with very high sensitivity, making it ideal for clinical diagnostic testing of FFPEs and other mixed samples.

[0096] Examples of sequencing techniques, often referred to as Next Generation Sequencing (NGS) techniques include, but are not limited to Sequencing by Synthesis (SBS), Massively Parallel Signature Sequencing (MPSS), Polony sequencing, pyrosequencing, Reversible dye-terminator sequencing, SOLiD sequencing, Ion semiconductor sequencing, DNA nanoball sequencing, Helioscope single molecule sequencing, Single molecule real time (SMRT) sequencing, Single molecule real time (RNAP) sequencing, and Nanopore DNA sequencing.

[0097] MPSS was a bead-based method that used a complex approach of adapter ligation followed by adapter decoding, reading the sequence in increments of four nucleotides; this method made it susceptible to sequence-specific bias or loss of specific sequences.

[0098] Polony sequencing, combined an in vitro paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an E. coli genome at an accuracy of > 99.9999% and a cost approximately 1/10 that of Sanger sequencing.

[0099] A parallelized version of pyrosequencing, the method amplifies DNA inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony. The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence read-outs. This technology provides intermediate read length and price per base compared to Sanger sequencing on one end and Solexa and SOLiD on the other.

[00100] SBS is a sequencing technology based on reversible dye-terminators. DNA molecules are first attached to primers on a flowcell and amplified so that local clonal colonies are formed. Four types of reversible terminator bases (RT-bases) are added, and non-incorporated nucleotides are washed away. Unlike pyrosequencing, the DNA can only be extended one nucleotide at a time. A camera takes images of the fluorescently labeled nucleotides, then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing the next cycle.

[00101] SOLiD technology employs sequencing by ligation. Here, a pool of all possible oligonucleotides of a fixed length are labeled according to the sequenced position.

[00102] Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal informative of the nucleotide at that position. Before sequencing, the DNA is amplified by emulsion PCR. The resulting bead, each containing only copies of the same DNA molecule, are deposited on a glass slide. The result is sequences of quantities and lengths comparable to Illumina sequencing.

[00103] Ion semiconductor sequencing is based on using standard sequencing chemistry, but with a novel, semiconductor based detection system. This method of sequencing is based on the detection of hydrogen ions that are released during the polymerization of DNA, as opposed to the optical methods used in other sequencing systems. A micro well containing a template DNA strand to be sequenced is flooded with a single type

of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal.

[00104] DNA nanoball sequencing is a type of high throughput sequencing technology used to determine the entire genomic sequence of an organism. The method uses rolling circle replication to amplify small fragments of genomic DNA into DNA nanoballs. Unchained sequencing by ligation is then used to determine the nucleotide sequence. This method of DNA sequencing allows large numbers of DNA nanoballs to be sequenced per run.

[00105] Helicos Biosciences Corporation's single-molecule sequencing uses DNA fragments with added poly A tail adapters, which are attached to the flow cell surface. The next steps involve extension-based sequencing with cyclic washes of the flow cell with fluorescently labeled nucleotides (one nucleotide type at a time, as with the Sanger method). The reads are performed by the Helioscope sequencer.

[00106] Single molecule real time (SMRT) sequencing is based on the SBS approach. The DNA is synthesized in zero-mode wave-guides (ZMWs) - small well-like containers with the capturing tools located at the bottom of the well. The sequencing is performed with use of unmodified polymerase (attached to the ZMW bottom) and fluorescently labeled nucleotides flowing freely in the solution. The wells are constructed in a way that only the fluorescence occurring by the bottom of the well is detected. The fluorescent label is detached from the nucleotide at its incorporation into the DNA strand, leaving an unmodified DNA strand.

[00107] Single molecule real time sequencing based on RNA polymerase (RNAP), which is attached to a polystyrene bead, with distal end of sequenced DNA is attached to another bead, with both beads being placed in optical traps. RNAP motion during transcription brings the beads in closer and their relative distance changes, which can then be recorded at a single nucleotide resolution. The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types (similarly to Sangers method).

[00108] Nanopore sequencing is based on the readout of electrical signal occurring at nucleotides passing by alpha-hemolysin pores covalently bound with cyclodextrin. The DNA passing through the nanopore changes its ion current. This change is dependent on the shape, size and length of the DNA sequence. Each type of the nucleotide blocks the ion flow through the pore for a different period of time.

[00109] VisiGen Biotechnologies uses a specially engineered DNA polymerase. This polymerase acts as a sensor - having incorporated a donor fluorescent dye by its active centre. This donor dye acts by FRET (fluorescent resonant energy transfer), inducing fluorescence of differently labeled nucleotides. This approach allows reads performed at the speed at which polymerase incorporates nucleotides into the sequence (several hundred per second). The nucleotide fluorochrome is released after the incorporation into the DNA strand.

[00110] Mass spectrometry may be used to determine mass differences between DNA fragments produced in chain-termination reactions.

[00111] SBS technology is capable of overcoming the limitations of existing pyrosequencing based NGS platforms.

[00112] Such technologies rely on complex enzymatic cascades for read out, are unreliable for the accurate determination of the number of nucleotides in homopolymeric regions and require excessive amounts of time to run individual nucleotides across growing DNA strands. The SBS NGS platform uses a direct sequencing approach to produce a sequencing strategy with very a high precision, rapid pace and low cost.

[00113] One exemplary SBS sequencing is initialized by fragmenting of the template DNA into fragments, amplification, annealing of DNA sequencing primers, and, for example, finally affixing as a high-density array of spots onto a glass chip. The array of DNA fragments are sequenced by extending each fragment with modified nucleotides containing cleavable chemical moieties linked to fluorescent dyes capable of discriminating all four possible nucleotides. The array is scanned continuously by a high-resolution electronic camera (Measure) to determine the fluorescent intensity of each base (A, C, G or T) that was newly incorporated into the extended DNA fragment. After the incorporation of each modified base the array is exposed to cleavage chemistry to break off the fluorescent dye and end cap allowing additional bases to be added. The process is then repeated until the fragment is completely sequenced or maximal read length has been achieved.

[00114] In another aspect, real-time PCR is used in detecting gene mutations, including for example but not limited to SNPs. In some embodiments, detection of SNPs in specific gene candidates is performed using real-time PCR, based on the use of intramolecular quenching of a fluorescent molecule by use of a tethered quenching moiety. Thus, according to exemplary embodiments, real-time PCR methods also include the use of molecular beacon technology. The molecular beacon technology utilizes hairpin-shaped molecules with an internally-quenched fluorophore whose fluorescence is restored by binding to a DNA target of interest (See, *e.g.*, Kramer, R. *et al. Nat. Biotechnol.* 14:303-308, 1996). In some embodiments, increased binding of the molecular beacon probe to the accumulating PCR product is used to specifically detect SNPs present in genomic DNA.

[00115] For the design of Real-Time PCR assays, several parts are coordinated, including the DNA fragment that is flanked by the two primers and subsequently amplified, often referred to as the amplicon, the two primers and the detection probe or probes to be used.

[00116] In some embodiments, a SNP site in a sample from the subject may be amplified by the amplification methods described herein or any other amplification methods known in the art. The nucleic acids in a sample may or may not be amplified prior to contacting the SNP site with a probe described herein, using a universal amplification method (*e.g.*, whole genome amplification and whole genome PCR).

[00117] Real-time PCR relies on the visual emission of fluorescent dyes conjugated to short polynucleotides (termed “detection probes”) that associate with genomic alleles in a sequence-specific fashion or on fluorescent molecules that intercalate into double stranded DNA referred to as quantitative or qPCR. Real-time PCR probes differing by a single nucleotide can be differentiated in a real-time PCR assay by the conjugation and detection of probes that fluoresce at different wavelengths. Real-Time PCR finds use in detection applications (diagnostic applications), quantification applications and genotyping applications.

[00118] Several related methods for performing real-time PCR are disclosed in the art, including assays that rely on TAQMAN® probes (U.S. Pat. Nos. 5,210,015 and 5,487,972, and Lee *et al.*, *Nucleic Acids Res.* 21:3761-6, 1993), molecular beacon probes (U.S. Pat. Nos. 5,925,517 and 6,103,476, and Tyagi and Kramer, *Nat. Biotechnol.* 14:303-8, 1996), self-probing amplicons (scorpions) (U.S. Pat. No. 6,326,145, and Whitcombe *et al.*, *Nat.*

Biotechnol. 17:804-7, 1999), Amplisensor (Chen *et al.*, *Appl. Environ. Microbiol.* 64:4210-6, 1998), Amplifluor (U.S. Pat. No. 6,117,635, and Nazarenko *et al.*, *Nucleic Acids Res.* 25:2516-21, 1997, displacement hybridization probes (Li *et al.*, *Nucleic Acids Res.* 30:E5, 2002), DzyNA-PCR (Todd *et al.*, *Clin. Chem.* 46:625-30, 2000), fluorescent restriction enzyme detection (Cairns *et al.*, *Biochem. Biophys. Res. Commun.* 318:684-90, 2004) and adjacent hybridization probes (U.S. Pat. No. 6,174,670 and Wittwer *et al.*, *Biotechniques* 22:130-1, 134-8, 1997).

[00119] One of the many suitable genotyping procedures is the TAQMAN® allelic discrimination assay. In some instances of this assay, an oligonucleotide probe labeled with a fluorescent reporter dye at the 5' end of the probe and a quencher dye at the 3' end of the probe is utilized. The proximity of the quencher to the intact probe maintains a low fluorescence for the reporter. During the PCR reaction, the 5' nuclease activity of DNA polymerase cleaves the probe, and separates the dye and quencher. This results in an increase in fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The 5' nuclease activity of DNA polymerase cleaves the probe between the reporter and the quencher only if the probe hybridizes to the target and is amplified during PCR. The probe is designed to straddle a target SNP position and hybridize to the nucleic acid molecule only if a particular SNP allele is present.

[00120] Real-time PCR methods include a variety of steps or cycles as part of the methods for amplification. These cycles include denaturing double-stranded nucleic acids, annealing a forward primer, a reverse primer and a detection probe to the target genomic DNA sequence and synthesizing (*i.e.*, replicating) second-strand DNA from the annealed forward primer and the reverse primer. This three step process is referred to herein as a cycle.

[00121] In some embodiments, about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or 60 cycles are employed. In some embodiments, about 10 to about 60 cycles, about 20 to about 50 or about 30 to about 40 cycles are employed. In some embodiments, 40 cycles are employed.

[00122] In some embodiments, the denaturing double-stranded nucleic acids step occurs at a temperature of about 80°C to 100°C, about 85°C to about 99°C, about 90°C to about 95°C for about 1 second to about 5 seconds, about 2 seconds to about 5 seconds, or

about 3 seconds to about 4 seconds. In some embodiments, the denaturing double-stranded nucleic acids step occurs at a temperature of 95°C for about 3 seconds.

[00123] In some embodiments, the annealing a forward primer, a reverse primer and a detection probe to the target genomic DNA sequence step occurs at about 40°C to about 80°C, about 50°C to about 70°C, about 55°C to about 65°C for about 15 seconds to about 45 seconds, about 20 seconds to about 40 seconds, about 25 seconds to about 35 seconds. In some embodiments, the annealing a forward primer, a reverse primer and a detection probe to the target genomic DNA sequence step occurs at about 60°C for about 30 seconds.

[00124] In some embodiments, the synthesizing (*i.e.*, replicating) second-strand DNA from the annealed forward primer and the reverse primer occurs at about 40°C to about 80°C, about 50°C to about 70°C, about 55°C to about 65°C for about 15 seconds to about 45 seconds, about 20 seconds to about 40 seconds, about 25 seconds to about 35 seconds. In some embodiments, the annealing a forward primer, a reverse primer and a detection probe to the target genomic DNA sequence step occurs at about 60°C for about 30 seconds.

[00125] In some embodiments, it was found that about 1 µL, about 2 µL, about 3 µL, about 4 µL or about 5 µL of a genomic DNA sample prepared according to the present methods described herein, are combined with only about 0.05 µL, about 0.10 µL about 0.15 µL, about 0.20 µL, about 0.25 µL or about 0.25 µL of a 30X, 35X, 40X, 45X, 50X or 100X real-time PCR assay mix and distilled water to form the PCR master mix. In some embodiments, the PCR master mix has a final volume of about 5 µL, about 6 µL, about 7 µL, about 8 µL, about 9 µL, about 10 µL, about 11 µL, about 12 µL, about 13 µL, about 14 µL, about 15 µL, about 16 µL, about 17 µL, about 18 µL, about 19 µL or about 20 µL or more. In some embodiments, it was found that 2 µL of a genomic DNA sample prepared as described above, are combined with only about 0.15 µL of a 40X real-time PCR assay mix and 2.85 µL of distilled water in order to form the PCR master mix.

[00126] While exemplary reactions are described herein, one of skill would understand how to modify the temperatures and times based on the probe design. Moreover, the present methods contemplate any combination of the above times and temperatures.

[00127] In some embodiments, primers are tested and designed in a laboratory setting. In some embodiments, primers are designed by computer based *in silico* methods. Primer sequences are based on the sequence of the amplicon or target nucleic acid sequence that is to

be amplified. Shorter amplicons typically replicate more efficiently and lead to more efficient amplification as compared to longer amplicons.

[00128] In designing primers, one of skill would understand the need to take into account melting temperature (T_m ; the temperature at which half of the primer-target duplex is dissociated and becomes single stranded and is an indication of duplex stability; increased T_m indicates increased stability) based on GC and AT content of the primers being designed as well as secondary structure considerations (increased GC content can lead to increased secondary structure). T_m 's can be calculated using a variety of methods known in the art and those of skill would readily understand such various methods for calculating T_m ; such methods include for example but are not limited to those available in online tools such as the T_m calculators available on the World Wide Web at promega.com/techserv/tools/biomath/calc11.htm. Primer specificity is defined by its complete sequence in combination with the 3' end sequence, which is the portion elongated by Taq polymerase. In some embodiments, the 3' end should have at least 5 to 7 unique nucleotides not found anywhere else in the target sequence, in order to help reduce false-priming and creation of incorrect amplification products. Forward and reverse primers typically bind with similar efficiency to the target. In some instances, tools such as NCBI BLAST (located on the World Wide Web at ncbi.nlm.nih.gov) are employed to performed alignments and assist in primer design.

[00129] Those of skill in the art would be well aware of the basics regarding primer design for a target nucleic acid sequence and a variety of reference manuals and texts have extensive teachings on such methods, including for example, *Molecular Cloning* (three volume set, Cold Spring Harbor Laboratory Press, 2012) and *Current Protocols (Genetics and Genomics; Molecular Biology; 2003-2013)* and *Real-Time PCR in Microbiology: From Diagnostics to Characterization* (Ian M. MacKay, Calster Academic Press; 2007); PrimerAnalyser Java tool available on the World Wide Web at primerdigital.com/tools/PrimerAnalyser.html and Kalendar R, *et al.* (*Genomics*, 98(2): 137-144 (2011)), all of which are incorporated herein in their entireties for all purposes.

[00130] An additional aspect of primer design is primer complexity or linguistic sequence complexity (*see*, Kalendar R, *et al.* (*Genomics*, 98(2): 137-144 (2011))). Primers with greater linguistic sequence complexity (*e.g.*, nucleotide arrangement and composition) are typically more efficient. In some embodiments, the linguistic sequence complexity

calculation method is used to search for conserved regions between compared sequences for the detection of low-complexity regions including simple sequence repeats, imperfect direct or inverted repeats, polypurine and polypyrimidine triple-stranded cDNA structures, and four-stranded structures (such as G-quadruplexes). In some embodiments, linguistic complexity (LC) measurements are performed using the alphabet-capacity L-gram method (see, A. Gabrielian, A. Bolshoy, *Computer & Chemistry* 23:263–274 (1999) and Y.L. Orlov, V.N. Potapov, Complexity: an internet resource for analysis of DNA sequence complexity, *Nucleic Acids Res.* 32: W628–W633(2004)) along the whole sequence length and calculated as the sum of the observed range (ξ) from 1 to L size words in the sequence divided by the sum of the expected (E) value for this sequence length. Some G-rich (and C-rich) nucleic acid sequences fold into four-stranded DNA structures that contain stacks of G-quartets (see, the World Wide Web at quadruplex.org). In some instances, these quadruplexes are formed by the intermolecular association of two or four DNA molecules, dimerization of sequences that contain two G-bases, or by the intermolecular folding of a single strand containing four blocks of guanines (see, P.S. Ho, *PNAS*, 91:9549–9553 (1994); I.A. Il'icheva, V.L. Florent'ev, *Russian Journal of Molecular Biology* 26:512–531(1992); D. Sen, W. Gilbert, *Methods Enzymol.* 211:191–199 (1992); P.A. Rachwal, K.R. Fox, *Methods* 43:291–301 (2007); S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, K. Neidle, *Nucleic Acids Res.* 34:5402–5415 (2006); A. Guédin, J. Gros, P. Alberti, J. Mergny, *Nucleic Acids Res.* 38:7858–7868 (2010); O. Stegle, L. Payet, J.L. Mergny, D.J. MacKay, J.H. Leon, *Bioinformatics* 25:i374–i382 (2009); in some instances, these are eliminated from primer design because of their low linguistic complexity, LC=32% for (TTAGGG)₄.

[00131] These methods include various bioinformatics tools for pattern analysis in sequences having GC skew, $(G-C)/(G+C)$, AT skew, $(A-T)/(A+T)$, CG–AT skew, $(S-W)/(S+W)$, or purine–pyrimidine $(R-Y)/(R+Y)$ skew regarding CG content and melting temperature and provide tools for determining linguistic sequence complexity profiles. For example the GC skew in a sliding window of n, where n is a positive integer, bases is calculated with a step of one base, according to the formula, $(G-C)/(G+C)$, in which G is the total number of guanines and C is the total number of cytosines for all sequences in the windows (Y. Benita, *et al.*, *Nucleic Acids Res.* 31:e99 (2003)). Positive GC-skew values indicated an overabundance of G bases, whereas negative GC-skew values represented an overabundance of C bases. Similarly, other skews are calculated in the sequence. Such

methods, as well as others, are employed to determine primer complexity in some embodiments.

[00132] According to non-limiting example embodiments, real-time PCR is performed using exonuclease primers (TAQMAN® probes). In such embodiments, the primers utilize the 5' exonuclease activity of thermostable polymerases such as Taq to cleave dual-labeled probes present in the amplification reaction (*See, e.g., Wittwer, C. et al. Biotechniques 22:130-138, 1997*). While complementary to the PCR product, the primer probes used in this assay are distinct from the PCR primer and are dually-labeled with both a molecule capable of fluorescence and a molecule capable of quenching fluorescence. When the probes are intact, intramolecular quenching of the fluorescent signal within the DNA probe leads to little signal. When the fluorescent molecule is liberated by the exonuclease activity of Taq during amplification, the quenching is greatly reduced leading to increased fluorescent signal. Non-limiting examples of fluorescent probes include the 6-carboxy-fluorescein moiety and the like. Exemplary quenchers include Black Hole Quencher 1 moiety and the like.

[00133] A variety of PCR primers can find use with the disclosed methods. Exemplary primers include but are not limited to those described herein.

[00134] A variety of detection probes can find use with the disclosed methods and are employed for genotyping and or for quantification. Detection probes commonly employed by those of skill in the art include but are not limited to hydrolysis probes (also known as TAQMAN® probes, 5' nuclease probes or dual-labeled probes), hybridization probes, and Scorpion primers (which combine primer and detection probe in one molecule).

[00135] In some embodiments, detection probes contain various modifications. In some embodiments, detection probes include modified nucleic acid residues, such as but not limited to 2'-O-methyl ribonucleotide modifications, phosphorothioate backbone modifications, phosphorodithioate backbone modifications, phosphoramidate backbone modifications, methylphosphonate backbone modifications, 3' terminal phosphate modifications and/or 3' alkyl substitutions.

[00136] In some embodiments, the detection probe has increased affinity for a target sequence due to modifications. Such detection probes include detection probes with increased length, as well as detection probes containing chemical modifications. Such modifications include but are not limited to 2'-fluoro (2'-deoxy-2'-fluoro-nucleosides)

modifications, LNAs (locked nucleic acids), PNAs (peptide nucleic acids), ZNAs (zip nucleic acids), morpholinos, methylphosphonates, phosphoramidates, polycationic conjugates and 2'-pyrene modifications. In some embodiments, the detector probes contains one or more modifications including 2' fluoro modifications (aka, 2'-Deoxy-2'-fluoro-nucleosides), LNAs (locked nucleic acids), PNAs (peptide nucleic acids), ZNAs (zip nucleic acids), morpholinos, methylphosphonates, phosphoramidates, and/or polycationic conjugates.

[00137] In some embodiments, the detection probes contain detectable moieties, such as those described herein as well as any detectable moieties known to those of skill in the art. Such detectable moieties include for example but are not limited to fluorescent labels and chemiluminescent labels. Examples of such detectable moieties can also include members of FRET pairs. In some embodiments, the detection probe contains a detectable entity.

[00138] Examples of fluorescent labels include but are not limited to AMCA, DEAC (7-Diethylaminocoumarin-3-carboxylic acid); 7-Hydroxy-4-methylcoumarin-3; 7-Hydroxycoumarin-3; MCA (7-Methoxycoumarin-4-acetic acid); 7-Methoxycoumarin-3; AMF (4'-(Aminomethyl)fluorescein); 5-DTAF (5-(4,6-Dichlorotriazinyl)aminofluorescein); 6-DTAF (6-(4,6-Dichlorotriazinyl)aminofluorescein); 6-FAM (6-Carboxyfluorescein; aka FAM; including TAQMAN® FAM™); TAQMAN VIC®; 5(6)-FAM cadaverine; 5-FAM cadaverine; 5(6)-FAM ethylenediamme; 5-FAM ethylenediamme; 5-FITC (FITC Isomer I; fluorescein-5-isothiocyanate); 5-FITC cadaverin; Fluorescein-5-maleimide; 5-IAF (5-Iodoacetamidofluorescein); 6-JOE (6-Carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein); 5-CR110 (5-Carboxyrhodamine 110); 6-CR110 (6-Carboxyrhodamine 110); 5-CR6G (5-Carboxyrhodamine 6G); 6-CR6G (6-Carboxyrhodamine 6G); 5(6)-Carboxyrhodamine 6G cadaverine; 5(6)-Carboxyrhodamine 6G ethylenediamme; 5-ROX (5-Carboxy-X-rhodamine); 6-ROX (6-Carboxy-X-rhodamine); 5-TAMRA (5-Carboxytetramethylrhodamine); 6-TAMRA (6-Carboxytetramethylrhodamine); 5-TAMRA cadaverine; 6-TAMRA cadaverine; 5-TAMRA ethylenediamme; 6-TAMRA ethylenediamme; 5-TMR C6 maleimide; 6-TMR C6 maleimide; TR C2 maleimide; TR cadaverine; 5-TRITC; G isomer (Tetramethylrhodamine-5-isothiocyanate); 6-TRITC; R isomer (Tetramethylrhodamine-6-isothiocyanate); Dansyl cadaverine (5-Dimethylaminonaphthalene-1-(N-(5-aminopentyl))sulfonamide); EDANS C2 maleimide; fluorescamine; NBD; and pyrromethene and derivatives thereof.

[00139] Examples of chemiluminescent labels include but are not limited to those labels used with Southern Blot and Western Blot protocols (see, for *e.g.*, Sambrook and

Russell, Molecular Cloning: A Laboratory Manual, (3rd ed.) (2001); incorporated by reference herein in its entirety). Examples include but are not limited to -(2'-spiroadamantane)-4-methoxy-4-(3"-phosphoryloxy)phenyl-1,2-dioxetane (AMPPD); acridinium esters and adamantyl-stabilized 1,2-dioxetanes, and derivatives thereof.

[00140] The labeling of probes is known in the art. The labeled probes are used to hybridize within the amplified region during amplification. The probes are modified so as to avoid them from acting as primers for amplification. The detection probe is labeled with two fluorescent dyes, one capable of quenching the fluorescence of the other dye. One dye is attached to the 5' terminus of the probe and the other is attached to an internal site, so that quenching occurs when the probe is in a non-hybridized state.

[00141] Typically, real-time PCR probes consist of a pair of dyes (a reporter dye and an acceptor dye) that are involved in fluorescence resonance energy transfer (FRET), whereby the acceptor dye quenches the emission of the reporter dye. In general, the fluorescence-labeled probes increase the specificity of amplicon quantification.

[00142] Real-time PCR that are used in some embodiments of the disclosed methods also include the use of one or more hybridization probes (i.e., detection probes), as determined by those skilled in the art, in view of this disclosure. By way of non-limiting example, such hybridization probes include but are not limited to one or more of those provided in the described methods. Exemplary probes, such as the HEX channel and/or FAM channel probes, are understood by one skilled in the art.

[00143] According to example embodiments, detection probes and primers are conveniently selected *e.g.*, using an in silico analysis using primer design software and cross-referencing against the available nucleotide database of genes and genomes deposited at the National Center for Biotechnology Information (NCBI). Some additional guidelines may be used for selection of primers and/or probes in some embodiments. For example, in some embodiments, the primers and probes are selected such that they are close together, but not overlapping. In some embodiments, the primers may have the same (or close T_M) (*e.g.*, between about 58 °C and about 60 °C). In some embodiments, the T_M of the probe is approximately 10 °C higher than that selected for the T_M of the primers. In some embodiments, the length of the probes and primers is selected to be between about 17 and 39 base pairs, etc. These and other guidelines are used in some instances by those skilled in the art in selecting appropriate primers and/or probes.

[00144] In some embodiments, the SNP described herein may be detected by melting curve analysis using the detection probes above. For example, the melting curves of short oligonucleotide probes hybridized to a region containing the SNP of interest may be analyzed. Two probes are used in these reactions, each one being complimentary to a particular allele at the SNP in question. Perfectly matched probes are more stable and have a higher melting temperature compared to mismatched probes. Hence, SNP genotypes are inferred according to the characteristic melting curves produced by annealing and melting either matched or mismatched oligonucleotide probes.

[00145] In one aspect, the methods described herein may include detecting the two or more SNPs described herein by hybridizing at least one detection probe to a nucleotide molecule from a sample or its amplicons and detecting the at least one detection probe.

[00146] In another aspect, diagnostic testing is employed to determine one or more genetic conditions by detection of any of a variety of mutations. In some embodiments, diagnostic testing is used to confirm a diagnosis when a particular condition is suspected based on for example physical manifestations, signs and/or symptoms as well as family history information. In some embodiments, the results of a diagnostic test assist those of skill in the medical arts in determining an appropriate treatment regimen for a given subject and allow for more personalized and more effective treatment regimens. In some embodiments, a treatment regimen include any of a variety of pharmaceutical treatments, surgical treatments, lifestyles changes or a combination thereof as determined by one of skill in the art.

[00147] The nucleic acids obtained by the disclosed methods are useful in a variety of diagnostic tests, including tests for detecting mutations such as deletions, insertions, transversions and transitions. In some embodiments, such diagnostics are useful for identifying unaffected individuals who carry one copy of a gene for a disease that requires two copies for the disease to be expressed, identifying unaffected individuals who carry one copy of a gene for a disease in which the information could find use in developing a treatment regimen, preimplantation genetic diagnosis, prenatal diagnostic testing, newborn screening, genealogical DNA test (for genetic genealogy purposes), presymptomatic testing for predicting or diagnosing KC.

[00148] In some embodiments, newborns can be screened. In some embodiments, newborn screening includes any genetic screening employed just after birth in order to identify genetic disorders. In some embodiments, newborn screening finds use in the

identification of genetic disorders so that a treatment regimen is determined early in life. Such tests include but are not limited to testing infants for phenylketonuria and congenital hypothyroidism.

[00149] In some embodiments, carrier testing is employed to identify people who carry a single copy of a gene mutation. In some cases, when present in two copies, the mutation can cause a genetic disorder. In some cases, one copy is sufficient to cause a genetic disorder. In some cases, the presence of two copies is contra-indicated for a particular treatment regimen, such as the presence of the Avellino mutation and pre-screening prior to performing surgical procedures in order to ensure the appropriate treatment regimen is pursued for a given subject. In some embodiments, such information is also useful for individual contemplating procreation and assists individuals with making informed decisions as well as assisting those skilled in the medical arts in providing important advice to individual subjects as well as subjects' relatives.

[00150] In some embodiments, predictive and/or presymptomatic types of testing are used to detect gene mutations associated with a variety of disorders. In some cases, these tests are helpful to people who have a family member with a genetic disorder, but who may exhibit no features of the disorder at the time of testing. In some embodiments, predictive testing identifies mutations that increase a person's chances of developing disorders with a genetic basis, including for example but not limited to certain types of cancer. In some embodiments, presymptomatic testing is useful in determining whether a person will develop a genetic disorder, before any physical signs or symptoms appear. The results of predictive and presymptomatic testing provides information about a person's risk of developing a specific disorder and help with making decisions about an appropriate medical treatment regimen for a subject as well as for a subject's relatives. Predictive testing is also employed, in some embodiments, to detect mutations which are contra-indicated with certain treatment regimens, such as the presence of the Avellino mutation being contra-indicated with performing LASIK surgery and/or other refractive procedures, such as but not limited to Phototherapeutic keratectomy (PTK) and/or Photorefractive keratectomy (PRK). For example, subjects exhibiting the Avellino mutation should not undergo LASIK surgery or other refractive procedures. Similarly, in some cases, subjects with KC mutation(s) should not undergo LASIK surgery or other refractive procedures.

[00151] In some embodiments, diagnostic testing also includes pharmacogenomics which includes genetic testing that determines the influence of genetic variation on drug response. Information from such pharmacogenomic analyses finds use in determining and developing an appropriate treatment regimen. Those of skill in the medical arts employ information regarding the presence and/or absence of a genetic variation in designing appropriate treatment regimen.

[00152] In some embodiments, diseases whose genetic profiles are determined using the methods of the present disclosure include KC.

[00153] In some embodiments, the present methods find use in development of personalized medicine treatment regimens by providing the genomic DNA which is used in determining the genetic profile for an individual. In some embodiments, such genetic profile information is employed by those skilled in the art in order determine and/or develop a treatment regimen. In some embodiments, the presence and/or absence of various genetic variations and mutations identified in nucleic acids isolated by the described methods are used by those of skill in the art as part of a personalized medicine treatment regimen or plan. For example, in some embodiments, information obtained using the disclosed methods is compared to databases or other established information in order to determine a diagnosis for a specified disease and or determine a treatment regimen. In some cases, the information regarding the presence or absence of a genetic mutation in a particular subject is compared to a database or other standard source of information in order to make a determination regarding a proposed treatment regimen. In some cases, the presence of a genetic mutation indicates pursuing a particular treatment regimen. In some cases the absence of a genetic mutation indicates not pursuing a particular treatment regimen.

[00154] In some embodiments, information regarding the presence and/or absence of a particular genetic mutation is used to determine the treatment efficacy of treatment with the therapeutic entity, as well as to tailor treatment regimens for treatment with therapeutic entity. In some embodiments, information regarding the presence and/or absence of a genetic mutation is employed to determine whether to pursue a treatment regimen. In some embodiments, information regarding the presence and/or absence of a genetic mutation is employed to determine whether to continue a treatment regimen. In some embodiments, the presence and/or absence of a genetic mutation is employed to determine whether to discontinue a treatment regimen. In other embodiments, the presence and/or absence of a

genetic mutation is employed to determine whether to modify a treatment regimen. In some embodiments the presence and/or absence of a genetic mutation is used to determine whether to increase or decrease the dosage of a treatment that is being administered as part of a treatment regimen. In other embodiments, the presence and/or absence of a genetic mutation is used to determine whether to change the dosing frequency of a treatment administered as part of a treatment regimen. In some embodiments, the presence and/or absence of a genetic mutation is used to determine whether to change the number of dosages per day, per week, times per day of a treatment. In some embodiments the presence and/or absence of a genetic mutation is used to determine whether to change the dosage amount of a treatment. In some embodiments, the presence and/or absence of a genetic mutation is determined prior to initiating a treatment regimen and/or after a treatment regimen has begun. In some embodiments, the presence and/or absence of a genetic mutation is determined and compared to predetermined standard information regarding the presence or absence of a genetic mutation.

[00155] In some embodiments, a composite of the presence and/or absence of more than one genetic mutation is generated using the disclosed methods and such composite includes any collection of information regarding the presence and/or absence of more than one genetic mutation. In some embodiments, the presence or absence of 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 20 or more, 30 or more or 40 or more genetic mutations, for example, including those in Figures 1-5, is examined and used for generation of a composite. Exemplary information in some embodiments includes nucleic acid or protein information, or a combination of information regarding both nucleic acid and/or protein genetic mutations. Generally, the composite includes information regarding the presence and/or absence of a genetic mutation. In some embodiments, these composites are used for comparison with predetermined standard information in order to pursue, maintain or discontinue a treatment regimen.

[00156] In some embodiments, KC is predicted and/or detected for example through detection of the two or more genetic variants described herein, for example, including at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200 or 250 variants selected from but not limited to those listed in Figure 1.

[00157] The present disclosure also provides methods to assist with differential diagnosis. In some embodiments, KC is distinguished from pellucid marginal degeneration,

keratoglobus, contact lens induced corneal warpage, and/or corneal ectasia post excimer laser treatment through detection of the two or more genetic variants described herein, for example, including at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200 or 250 variants selected from but not limited to those listed in Figure 1.

[00158] In some embodiments, the two or more genetic variants are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200 or 250 variants selected from the group listed in Figure 2, and the subject is Afro-American.

[00159] In some embodiments, the two or more genetic variants are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200 or 250 variants selected from the group listed in Figure 3, and the subject is Caucasian.

[00160] In some embodiments, the two or more genetic variants are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200 or 250 variants selected from the group listed in Figure 4, and the subject is Hispanic.

[00161] In some embodiments, the two or more genetic variants are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200 or 250 variants selected from the group listed in Figure 5 and the subject is East Asian.

[00162] In some embodiments, the detection of two or more genetic variants is combined with a physical examination in order to diagnose KC or predict the risk of developing KC. Such a physical examination can include an eye examination as well as ancillary tests to assess corneal curvature, astigmatism and thickness. In some embodiments, the best potential vision of the subject is evaluated. Components of the eye exam can include but are not limited to medical history (including, for example, change in eye glass prescription, decreased vision, history of eye rubbing, medical problems, allergies, and/or sleep patterns); assessment of relevant aspects of the subject's mental and physical status; visual acuity with current correction (the power of the present correction recorded) at distance and when appropriate at near and far distances; measurement of best corrected visual acuity with spectacles and/or hard or gas permeable contact lenses (with refraction when indicated); measurement of pinhole visual acuity; external examination (lids, lashes, lacrimal apparatus, orbit); examination of ocular alignment and motility; assessment of pupillary function; measurement of intraocular pressure (IOP); slit-lamp biomicroscopy of the anterior

segment; dilated examination (including for example, dilated examination of the lens, macula, peripheral retina, optic nerve, and vitreous); and Keratometry/Computerized Topography/Computerized Tomography/Ultrasound Pachymetry.

[00163] In some embodiments, the detection of two or more genetic variants is in combination with one or more indications or signs of KC development in order to diagnose KC or predict the risk of developing KC. In some embodiments, the sign is an early sign of KC. In some embodiments, an early sign of KC includes but is not limited to asymmetric refractive error with high or progressive astigmatism; keratometry showing high astigmatism and irregularity (axis that do not add to 180 degrees); scissoring of the red reflex on ophthalmoscopy or retinoscopy; inferior steepening, skewed axis, or elevated keratometry values on K reading and computerized corneal topography; corneal thinning, especially in inferior cornea (maximum corneal thinning corresponds to the site of maximum steepening or prominence); Rizzuti's sign or a conical reflection on nasal cornea when a penlight is shone from the temporal side; Fleischer ring, an iron deposit often present within the epithelium around the base of the cone. It is brown in color and best visualized with a cobalt blue filter; or Vogt's striae, fine, roughly vertically parallel striations in the stroma (these generally disappear with firm pressure applied over the eyeball and re-appear when pressure is discontinued). In some embodiments, the sign is a late sign of KC. In some embodiments, a late sign of KC includes but is not limited to Munson's sign (a protrusion of the lower eyelid in downgaze); superficial scarring; break's in Bowman's membrane; acute hydrops (a condition where a break in Descemet's membrane allows aqueous fluid into the stroma causing severe corneal thickening, decreased vision and pain); or stromal scarring after resolution of acute hydrops (which paradoxically may improve vision in some cases by changing corneal curvature and reducing the irregular astigmatism).

[00164] In some embodiments, the detection of two or more genetic variants associated with an increased risk of developing KC can be used to assist with determining a treatment regimen for an individual suspected to have KC or predicted to develop KC in the future.

[00165] KC treatment regimens include a variety of treatment regimens directed to providing visual acuity and maintaining sight. Spectacles or soft toric contact lenses in mild cases can be used. Rigid gas permeable contact lenses are needed in the majority of cases to neutralize the irregular corneal astigmatism. The majority of subjects that can wear hard or gas-permeable contact lenses have a dramatic improvement in their vision. Specialty contact

lenses have been developed to better fit the irregular and steep corneas found in KC; these include (but not limited to) RoseK™, custom designed contact lenses (based on topography and/or wavefront measurements), semi-scleral contact lenses, piggy back lens use (soft and hard lens used at the same time), and scleral lenses. Subjects that become contact lens intolerant or do not have acceptable vision (e.g., from central scarring) proceed to surgical alternatives.

[00166] In some embodiments, the detection of two or more genetic variants as described herein can be used to begin an appropriate treatment early in an individual suspected to be a risk of developing KC. In some embodiments, treatments are directed to halting changes in the corneal shape. In some embodiments, the detection of two or more genetic variants that predict and increased risk of developing KC can allow for earlier and/or more frequent monitoring of the cornea in order to identify disease onset at an early stage. (i.e., identify early disease onset).

[00167] In some embodiments, treatment includes medical therapy for subjects who have an episode of corneal hydrops involves acute management of the pain and swelling. Subjects are usually given a cycloplegic agent, sodium chloride (Muro) 5% ointment and may be offered a pressure patch. After the pressure patch is removed subjects may still need to continue sodium chloride drops or ointment for several weeks to months until the episode of hydrops has resolved. Subjects are advised to avoid vigorous eye rubbing or trauma.

[00168] In another aspect, the detection of two or more SNPs as described herein can be used to begin early or regular monitoring in an individual suspected to be a risk of developing KC. In some embodiments, subjects can be followed on a 6-month to yearly basis to monitor the progression of the corneal-thinning and steepening and the resultant visual changes and to re-evaluate contact lens fit and care. In some embodiments, subjects who have developed hydrops are seen more frequently until the symptoms resolve.

[00169] In another aspect, the detection of two or more genetic variants as described herein can be used to diagnose KC in a subject. In some embodiments, after diagnosis, a treatment regimen includes surgical interventions. While initial treatment regimens focus on less invasive procedures, such as contact lens fitting if the subject does not exhibit corneal scarring. However, as subjects become intolerant or no longer benefit from contact lenses, surgery is the next option. Surgical options can include but are not limited to INTACS (i.e.,

implants, also known as ICRS or corneal rings), Anterior lamellar keratoplasty, or penetrating keratoplasty. Treatment can also include non- FDA approved treatments, which include but are not limited to the use of UV/riboflavin collagen cross-linking of the cornea to stiffen the cornea and possibly prevent progressive changes in shape and this treatment can be combined with excimer laser treatment, conductive keratoplasty, and/or INTACS. In some embodiments, surgeons can also use phakic intraocular lenses (IOLs) to address high myopia and some of the astigmatism.

[00170] In some embodiments, the surgical intervention includes intracorneal ring segments (INTACS; commercially available from Addition Technology), which have also been approved for the treatment of mild to moderate KC in subjects who are contact lens intolerant. In these cases, subjects must have a clear central cornea and a corneal thickness of > 450 microns where the segments are inserted, approximately at 7 mm optical zone. An advantage of INTACS is that they require no removal of corneal tissue, no intraocular incision, and leave the central cornea untouched. Most subjects will need spectacles and/or contact lenses post-operatively for best vision, but will have flatter corneas and easier use of lenses after the procedure. In some instances, INTACS can be removed and then other surgical options can be considered.

[00171] In some embodiments, the surgical intervention includes Anterior lamellary keratoplasty, which has resurfaced as an option for treating KC. It involves replacement of the central anterior cornea, leaving the subject's endothelium intact. The advantages are that the risk of endothelial graft rejection is eliminated, and there is less risk of traumatic rupture of the globe in the incision, since the endothelium and Descemet's and some stroma are left intact, and faster visual rehabilitation. There are several techniques including, deep anterior lamellar keratoplasty (DALK) and big bubble keratoplasty (BBK) to remove the anterior stroma, while leaving Descemet's layer and endothelium untouched. However, the procedures can be technically challenging requiring conversion to a penetrating keratoplasty, and post-operatively there is the possibility of interface haze leading to a decrease in best corrected visual acuity (BCVA); it is not clear if astigmatism is better treated with anterior vs penetrating keratoplasty. Penetrating keratoplasty has a high success rate and is the standard surgical treatment with a long track record of safety and efficacy. Risks of this procedure include infection and cornea rejection and risk of traumatic rupture at wound margin. Many subjects after penetrating keratoplasty (PK) may still need hard or gas-permeable contact lenses due to residual irregular astigmatism. Any type of refractive

procedure is considered a contraindication in keratoconic subjects due to the unpredictability of the outcome and risk of leading to increased and unstable irregular astigmatism.

[00172] Additionally, the treatment for keratoconus includes collagen cross-linking and corneal transplant. Collagen cross-linking is a new treatment that uses a special laser and eyedrops to promote “cross-linking” or strengthening of the collagen fibers that make up the cornea. This treatment may flatten or stiffen the cornea, preventing further protrusion. When good vision is no longer possible with other treatments, a corneal transplant may be recommended. In a corneal transplant, the diseased cornea is removed from your eye and is replaced it with a healthy donor cornea.

[00173] In one aspect, the disclosure provides methods for treating keratoconus in a subject, the method comprising diagnosing or prognosing KC and treating KC in the subject. In further embodiments, the treating may comprise wearing eye glasses or contact lenses, administering a cycloplegic agent, applying intracorneal ring segments, performing anterior lamellary keratoplasty, and/or performing collagen cross-linking or corneal transplant.

[00174] In another aspect, the disclosure provides a diagnostic kit for diagnosing, prognosing and/or treating KC. Any or all of the reagents described above may be packaged into a diagnostic kit. Such kits include any and/or all of the primers, probes, buffers and/or other reagents described herein in any combination. In some embodiments, the kit includes reagents for detection of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 variants selected from but not limited to those listed in Figure 1.

[00175] In some embodiments, the reagents in the kit are included as lyophilized powders. In some embodiments, the reagents in the kit are included as lyophilized powders with instructions for reconstitution. In some embodiments, the reagents in the kit are included as liquids. In some embodiments, the reagents are included in plastic and/or glass vials or other appropriate containers. In some embodiments the primers and probes are all contained in individual containers in the kit. In some embodiments, the primers are packaged together in one container, and the probes are packaged together in another container. In some embodiments, the primers and probes are packaged together in a single container.

[00176] In some embodiments, the kit further includes control gDNA and/or DNA samples. In some embodiments, the control DNA sample is normal (e.g., from a subject who

does not have KC). In some embodiments, the control DNA sample corresponds to the mutation being detected, including any of variants selected from the group listed in Figure 1.

[00177] In some embodiments, the concentration of the control DNA sample is 5 ng/ μ L, 10 ng/ μ L, 20 ng/ μ L, 30 ng/ μ L, 40 ng/ μ L, 50 ng/ μ L, 60 ng/ μ L, 70 ng/ μ L, 80 ng/ μ L, 90 ng/ μ L, 100 ng/ μ L, 110 ng/ μ L, 120 ng/ μ L, 130 ng/ μ L, 140 ng/ μ L, 150 ng/ μ L, 160 ng/ μ L, 170 ng/ μ L, 180 ng/ μ L, 190 ng/ μ L or 200 ng/ μ L. In some embodiments, the concentration of the control DNA sample is 50 ng/ μ L, 100 ng/ μ L, 150 ng/ μ L or 200 ng/ μ L. In some embodiments, the concentration of the control DNA sample is 100 ng/ μ L. In some embodiments, the control DNA samples have the same concentration. In some embodiments, the control DNA samples have different concentrations.

[00178] In some embodiments, the kit can further include buffers, for example, GTXpress TAQMAN® reagent mixture, or any equivalent buffer. In some embodiments, the buffer includes any buffer described herein.

[00179] In some embodiments, the kit can further include reagents for use in cloning, such as vectors (including, *e.g.*, M13 vector).

[00180] In some embodiments, the kit further includes reagents for use in purification of DNA.

[00181] In some embodiments, the kit further includes instructions for using the kit for the detection of corneal dystrophy in a subject. In some embodiments, these instructions include various aspects of the protocols described herein.

The Pilot Study

[00182] A study cohort consisted of 219 cases and 60 controls. The Caucasian group consisted of 70 individual cases and 33 family cases for a total of 104 cases plus 38 controls; the East Asian cohort consisted of 70 individual cases and 5 family cases for a total of 75 cases plus 20 controls; the Hispanic group consisted of 13 individual cases and 5 family cases for a total of 18 cases plus 1 control; the African-American group consisted of 15 individual cases and 3 family cases for a total of 18 cases; and the South Asian group consisted of 3 individual cases and 2 family cases totaling 5 cases and 1 control (Figures 1-6).

[00183] Samples and controls were collected from clinics in the USA, Canada, Czech Republic, Greece, Brazil, Northern Ireland, South Korea, and Mexico. Controls were

collected from individuals with neither a personal nor family history of eye diseases. Each clinic utilized criteria to diagnose KC based on the CLEK study and on a global consensus study on KC and ectatic diseases. In summary, all subjects underwent a comprehensive ophthalmological examination, and a diagnosis of KC was based on corneal topography with Placido-disk based reflection, corneal tomography and clinical findings on slit lamp examination. Corneal topography and pachymetry, mean keratometric value (K), steep K, maximum K, thinnest corneal thickness and central corneal thickness were measured utilizing a Scheimpflug camera system such as an Oculyzer II (Alcon Surgical, Ft. Worth, Tex., USA) or a Pentacam® HR (OCULUS Optikgeräte GmbH, Wetzlar, Germany). At some locations a Schwind Sirius (Schwind Eye-Tech Solutions, Kleinostheim, Germany) was also utilized to diagnose for high order aberrations (HOAs). If a family history was known, it was disclosed by the patient at the time of sample collection.

[00184] Sample collections were carried out with iSWAB collection kits (Mawi DNA Technologies, Hayward, CA, USA). In brief, collection kits contained 4 buccal swabs and a 1 mL solution containing an undisclosed preservative in a specialized 1.5 mL Eppendorf tube. Patients were required to rub the inner cheek with each of the 4 swabs collecting enough epithelial tissue to ensure a DNA yield of between 0.5 and 3.0 µg of genomic DNA. Each of the swabs was placed into an Eppendorf tube that is designed to scrape the collected cells from each of the buccal swabs. The tubes containing the collected epithelial cells were stored at 4°C until ready for use.

[00185] QIAamp® DNA blood mini kits from QIAGEN Inc. (Hilden, Germany) were used to carry out genomic DNA extractions. The DNA extraction protocol recommended for whole blood was utilized for all samples, and DNA was eluted from spin columns in 150 µl elution buffer provided in the kit. A concentration of 3.4 ng/µl was the minimum acceptable DNA concentration to yield at least 0.5 µg, the minimum needed for the WES library preparation.

[00186] The ACE Platform™ (Personalis Inc., Menlo Park, CA) was utilized for all whole exome sequencing (WES) runs, which were conducted by Personalis on an Illumina HiSeq 2000. The whole exome ACE Platform™ provides augmented coverage to regions outside the exome including regulatory regions for over 8,000 genes. Resulting sequence data was processed by Personalis, and variant call format (VCF) files were generated for all

cases and controls. Each VCF file consisted of approximately 150,000 variants found within the approximately 22,000 genes that make up the human exome.

[00187] VCFs were processed using BCFtools version 1.3.1 to left-align and normalize indels, split multi-allelic sites into multiple calls, and to check that reference bases matched the known reference (1000 Genomes Phase 1 and 3 GRCh37 reference). VCFtools version 0.1.15 was then used to purge all reference base calls and variants called on contigs outside of chr1-22XY. BCFtools version 1.3.1 was then used to merge all samples into a single variant 'database', from which samples across each ethnic group were extracted into sub-groups. Each sub-group was then converted to PLINK format and allele tallies for each variant counted for cases and controls using PLINK v1.90b3.38. PLINK results files were then modified using custom BASH scripts, with all variants then annotated using ANNOVAR.

[00188] Variants were annotated with three different scoring systems in order to determine the level of conservation in the region surrounding each variant. These were GERP++ where scores range from -12.3 to 6.17, with 6.17 being the most conserved, PhyloP which calculates a score based on 40+ genome alignments including both vertebrates and mammals, and SiPhy which utilizes 29 genome alignments (mammals) and produces a log odds ratio, with the higher value indicating higher conservation. Additional filtering was based on a minor allele frequency (MAF) of ≤ 0.05 or NA as documented in the Exome Aggregation Consortium (ExAC, <http://exac.broadinstitute.org/>), which contains data from 60,706 unrelated individuals. ExAc sub-populations were matched to the sample ethnic groups as follows: ExAc AFR (African/African-Americans), African-Americans; ExAc NFE (non-Finnish Europeans), Caucasians; ExAc EAS (East Asians), East Asians; and ExAc AMR (Hispanic (ad-mixed Americans), Hispanics.

[00189] In order to select variants most likely to be damaging and thus related to disease, the following criteria were applied: variants classified as missense, STOP gain/loss, nonsense, or frameshift/non-frameshift InDels were focused. These variants were further filtered to those within genes related to the cornea or KC, key terms through gene set enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery. The functional annotation chart tool was used with default categories plus 'GAD_Disease' and 'GAD_Disease_Class,' and a list of all enriched terms was derived with gene count 1 and EASE 1.0. Finally, pathology for each variant was gauged on the *in silico* predictions from 7 published methods: SIFT, PolyPhen 2 HDIV, PolyPhen 2 Hvar, LRT,

MutationTaster, MutationAssessor, and FATHMM. Each tool aims to determine the likely impact on the transcribed amino acid sequence and translated protein due to a missense change in the exonic DNA sequence, with each taking into account different metrics when arriving at a prediction. A variant would be classified as 100% pathogenic if it satisfied the following predictions from each tool: SIFT, *deleterious*; PolyPhen 2 HDIV, *probably damaging/possibly damaging*; PolyPhen 2 HVar, *probably damaging/possibly damaging*; LRT, *deleterious*; MutationTaster, *disease-causing-automatic, disease-causing*; MutationAssessor, *high*; FATHMM, *deleterious*. Variants classified as benign and/or common were only considered if relevant to the disease profile and present within the case samples at a higher MAF level than what is documented in ExAC. For this study, a common variant is defined as having an MAF within ExAC greater than 1% i.e., $\text{MAF (ExAC_All)} > 0.01$

[00190] For profiling the ethnicity of each sample, publicly-available 1000 Genomes Phase 3 VCF data (The 1000 Genomes Project Consortium, 2015) ($n=2,504$), available on a chromosome by chromosome basis was used. Samples within the study cohort were compared against these data. Individual 1000 Genomes chromosome VCFs were normalized as per the KC samples. All proceeding analyses were then conducted using PLINK v1.90b3.38.

[00191] Normalized VCFs were converted to PLINK format and then only matching variants based on dbSNP rs numbers across 1000 Genomes and keratoconus samples were retained. Variants were pruned from each 1000 Genomes chromosome and the keratoconus dataset based on the following parameters: only retain variants with $\text{MAF} > 0.2$ and those not under linkage disequilibrium based on: window size, 50; step size (variant count), 5; variance inflation factor (VIF) threshold, 1.5. Multi-allelic variants were further removed. All 1000 Genomes chromosomes and the KC dataset were then merged into a single project. A principal components analysis (PCA) was then performed. The sample eigenvalues were plotted for the first 3 principal components using R version 3.2.5 (2016-04-14). The 1000 Genomes were categorized into their respective super populations, i.e., African/African American, Hispanic (ad-mixed American), East Asian, European, and South Asian.

[00192] To predict the ethnicity of KC samples where ethnicity is not mentioned in the clinical record provided by the clinic, a simple multinomial logistic regression model was built using the 1000 Genomes data filtered as above. In this model, 1000 Genomes super

populations were the outcome and the first 20 principal components were predictors. The best principal components predicting the super populations were selected using forward-backward stepwise regression analysis and the Bayesian information criterion (BIC). This model achieved an area under the curve (AUC) of 0.987 (95% CI: 0.982-0.992) through receiver operating characteristics (ROC) analysis. This model was then used to predict the ethnicity of all 1000 Genomes and KC samples and plotted their respective prediction values. In all but one case, the predicted ethnicity of the KC sample matched the assumed ethnicity based on origin of shipment of the sample. All modeling was performed using R.

[00193] A relative risk (RR) score was created for the purpose of assigning a quantitative value for disease prediction for a subset of variants found within genes directly related to corneal structure and function. (Figure 7) The following steps were used in the calculation of risk scores. A Bayesian logistic regression model was first constructed with case/control status as outcome and variants selected for downstream analysis as predictors. The PhyloP conservation scores were supplied on the log odds scale for each variant used as a predictor in the model, with the mean prior being the mean PhyloP score across all variants called in each respective ethnic sub-group being analyzed. This model therefore produced an odds ratio (OR) for each variant that took relative allele tallies across cases and controls into account, and also the conservation of the region in which the variant was identified ('conservation ORs'), such that: greater conservation resulted in an increase in the OR; lower conservation resulted in a decrease. Risk scores were then directly calculated from the conservation ORs through multiplication by the number of *in silico* tools predicting a damaging outcome by the defined criteria previously mentioned. The risk scores for indels were left as their respective conservation ORs as the current *in silico* tools cannot provide predictions for these. Explained keratoconus variation is defined by McFadden's R².

[00194] The heterogeneity of the WES data and the establishment of ethnic subgroups: The study cohort consisted of 5 ethnic groups: Caucasians, East Asians, Hispanics, African Americans and South Asians. Given the ethnic diversity of this study, and in light of the known variations in the incidence and prevalence of KC across ethnic groups, how ethnicity might influence the genetic profile of the study group was determined. A PCA bi-plot was used to graph the entire KC cohort against 1000 Genomes Phase 3 VCF data; the sample cohort was segregated into sub-groups based on population variant patterns that occur naturally.

[00195] Genetic variants were identified over the entire exome with varying frequencies within each of the ethnic groups. A total of 1,117 variants located in 259 genes known to be associated with both syndromic and nonsyndromic eye disease were identified within the study cohort (Figure 1). Variants are defined here as missense single nucleotide polymorphisms (SNPs) and coding insertions and deletions (indels) predicted to alter protein function. Variants classified as benign were included if present within case samples at a higher minor allele frequency (MAF) than what is documented in the Exome Aggregation Consortium (ExAC).

[00196] Genes or the loci where genes are located were identified as relevant to the disease. For example, the results support that the common variants within the *ZNF469* gene, such as rs3812954, play a role in the etiology of KC. This variant was present within all case samples at a rate of 18.3% (Table 3), 25.5% among the Caucasian cohort and 18.1% within the East Asian cohort. Many of these types of variants were shared between ethnicities.

[00197] A common variant, rs6138482 within the *VSX1* gene was present in the study cohort at a MAF of 20.8%. Found in three of the ethnic groups, Caucasian, Hispanic and African American, it was most prevalent in the Caucasian group at 33.3% or 34 out of 102 cases. When considering the presence of any variant within an individual genome, the genotype, whether the variant is found in a heterozygous or homozygous form must also be taken into account.

[00198] Rare variants and the provision of a risk scoring strategy to predict pathology: Most variants were found in a specific individual, i.e., private variants; consequently, traditional statistical methodologies typically applied to GWAS and common variants failed to provide significance to the heterogenetic model that the data presented us. Given the extent of the findings over such a broad range of genes, an in-depth analysis was conducted on genes related to the structure and the function of the cornea. Since KC is a disease whose phenotype affects the cornea, quantifying a risk factor for variants within genes of the cornea is used, in some embodiments, as a diagnostic measure in a clinical setting.

[00199] In order to assess significance to a group of variants, a method was created that assigned a risk factor, which functioned to predict the pathology of the chosen variants. For this analysis, a total of 199 variants within 48 genes (Figure 7) related to corneal structure and function are represented.

[00200] Figure 7 lists an OR adjusted to the conservation for the region on the genome for each variant. The sensitivity and AUC based on the ROC for this set of variants informs us that for the Caucasian group (103 case samples) the panel successfully identified variants 95% of the time.

[00201] Genetic Testing: Furthermore, this work supports genetic testing for pre-symptomatic individuals who may be at risk due to family history or who are candidates for refractive surgery. Understanding the risk of developing KC before the symptoms appear will help to ensure proper diagnosis and treatment and would help to alleviate the trauma of physical discomfort and vision loss that this disease brings. A quantitative risk score can be used to assess the pathology of rare variants within genes related to the structure and function of the cornea. This model can be expanded to include other rare variants and even common variants. Variants were conservatively chosen to be used in demonstrating this tool, as the study cohort was limited in numbers, and it must be emphasized that the risk scores are relative to the sample set from which they are taken.

[00202] Table 1: Average number of rare variants from cornea genes present in case samples

Ethnicity	Ave. Variant Number	St. Dev.
Caucasian	2.17	2.53
East Asian	3.43	2.63
Hispanic	4.78	3.86
African American	11.11	5.17

[00203] Rare variants related to corneal structure and function were drawn from a larger list of variants found within the study cohort. Variants were further selected based on their presence in 1 or more case samples and 0 in ethnic-matched controls. The average variant count ranged from 2 to 5 variants per case with the exception of the African American cohort (Table 1).

[00204] In some embodiments, a higher order risk plot based on 3 or more variants within the genome of the patient is used as a predictor for KC.

WHAT IS CLAIMED:

1. A method for diagnosing or prognosing KC in a subject, the method comprising detecting two or more genetic variants in a sample from a subject, wherein the two or more genetic variants are selected from the group listed in Figure 1, and wherein the presence of two or more genetic variants is indicative of a diagnosis or prognosis of KC in the subject.
2. The method according to claim 1, wherein said variant detection is by a sequencing method.
3. The method according to claim 1 or 2, wherein the two or more genetic variants are selected from the group listed in Figure 2 and the subject is Afro-American.
4. The method according to claim 1 or 2, wherein two or more genetic variants are selected from the group listed in Figure 3 and the subject is Caucasian.
5. The method according to claim 1 or 2, wherein the two or more genetic variants are selected from the group listed in Figure 4 and the subject is Hispanic.
6. The method according to claim 1 or 2, wherein the two or more genetic variants are selected from the group listed in Figure 5 and the subject is East Asian.
7. The method according to any of the preceding claims, further comprising amplifying a nucleotide molecule from the sample from the subject.
8. The method according to any of the preceding claims, wherein the detecting comprises detecting the two or more genetic variants in a nucleotide molecule from the sample from the subject or its amplicons.
9. A method for predicting risk of developing KC in a subject, the method comprising detecting two or more genetic variants in a sample from a subject, wherein two or more genetic variants are selected from the group listed in Figure 1, and wherein the presence of two or more genetic variants is indicative of a risk for developing KC in the subject.

10. The method according to claim 9, wherein the two or more genetic variants are selected from the group listed in Figure 2 and the subject is Afro-American.
11. The method according to claim 9, wherein the two or more genetic variants are selected from the group listed in Figure 3 and the subject is Caucasian.
12. The method according to claim 9, wherein the two or more genetic variants are selected from the group listed in Figure 4 and the subject is Hispanic.
13. The method according to claim 9, wherein the two or more genetic variants are selected from the group listed in Figure 5 and the subject is East Asian.
14. The method according to any one of claims 9-13, further comprising amplifying a nucleotide molecule from the sample from the subject.
15. The method according to any one of claims 9-14, wherein the detecting comprises detecting the two or more genetic variants in a nucleotide molecule from the sample from the subject or its amplicons.
16. A method for developing a treatment regimen for the treatment of KC in a subject, the method comprising detecting two or more genetic variants in a sample from a subject, wherein the two or more genetic variants are selected from the group in Figure 1, and wherein the presence of two or more genetic variants is indicative of the need for a KC treatment regimen in the subject.
17. A method for treating KC in a subject, the method comprising detecting two or more genetic variants in a sample from a subject, wherein the two or more genetic variants are selected from the group listed in Figure 1 and treating KC in the subject.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

8/14

Figure 4

Chr	Position	Ref	Alt	Gene	Ref Change (Ref Gene)	RefSeq ID	MAF (FAC 10)	Control (FAC 10)	MAF (Control)	Human (FAC 10)	MAF (Human)
2	1882462	C	T	P2RX1	NM_012250:c.521C>A p.A169V	rs32111877	0.007	0	0.000000	1	0.007778
2	1882466	C	T	P2RX1	NM_012250:c.522A>G p.R169G	rs375388362	0.004	0	0.000000	1	0.007778
2	11535593	C	G	ADA	NM_000750:c.533G>A p.G168A	rs145946828	0.001	0	0.000000	1	0.007778
2	128930126	G	A	COX5A2	NM_000649:c.376T>A p.A1256V	rs59464605	0.008	0	0.000000	1	0.007778
2	125172623	A	G	PRPF8	NM_005340:c.518G>C p.A173V	rs1584251	0.008	0	0.000000	2	0.007778
2	115738098	G	C	SRRT5	NM_005222:c.488T>A p.A163P	rs76327971	0.003	0	0.000000	1	0.007778
2	227493381	T	C	COX4A4	NM_000521:c.417G>G p.S194G	rs3896949	0.003	0	0.000000	1	0.007778
2	228321213	A	G	COX4A4	NM_000521:c.417C>G p.S194V	rs18822264	0.002	0	0.000000	2	0.007778
2	228355763	C	A	COX4A4	NM_000521:c.464A>G p.D154E	rs7399514	0.002	0	0.000000	1	0.007778
3	19570912	S	T	NR1A2	NM_008406:c.1105G>A p.L369V	rs553518670	0.001	0	0.000000	1	0.007778
3	195513773	A	G	NR1A2	NM_008406:c.1337G>C p.L445P	rs155511263	0.002	0	0.000000	1	0.007778
5	142038441	T	C	DISA1	NM_001079121:c.423A>G p.A141V	rs77842435	0.001	0	0.000000	1	0.007778
5	142038961	T	C	DISA1	NM_001079121:c.423T>G p.A141V	rs3705870	0.001	1	0.000000	3	0.007778
7	20467132	T	C	ABCB9	NM_001165941:c.1120G>A p.A373V	rs202093881	0.007	0	0.000000	1	0.007778
7	20469233	C	T	ABCB9	NM_001165941:c.1377G>T p.T459I	rs207851280	0.005	0	0.000000	1	0.007778
7	207851281	CA	-	ABCB9	NM_001165941:c.1377G>T p.T459I	rs77168360	0.003	0	0.000000	1	0.007778
7	105592825	S	A	PRKG2	NM_001252428:c.588A>G p.G193E	rs78514317	0.005	0	0.000000	1	0.007778
9	35761050	T	C	TGN1	NM_000288:c.433G>A p.A143V	rs75515122	0.001	0	0.000000	1	0.007778
9	124081553	C	T	GBN	NM_000177:c.258A>G p.T86A	rs144414047	0.007	0	0.000000	1	0.007778
9	157767473	G	A	COL1A1	NM_000288:c.940G>A p.A313V	rs147398175	0.004	0	0.000000	1	0.007778
10	95418717	A	C	PDGEC	NM_000541:c.456G>C p.G152V	rs12651111	0.007	0	0.000000	1	0.007778
10	115119574	T	C	ZEB1	NM_001148091:c.123G>A p.L41V	rs3577180	0.002	0	0.000000	1	0.007778
10	33214867	S	A	ITGB1	NM_001088:c.718T>A p.R239K	rs43469491	0.003	0	0.000000	1	0.007778
12	59146577	S	A	KRT3	NM_001088:c.250T>G p.S84W	rs2517288	0.002	0	0.000000	1	0.007778
12	59146577	S	A	KRT3	NM_001088:c.250T>G p.S84W	rs2517288	0.004	0	0.000000	1	0.007778
15	86897921	C	T	ABCE1	NM_001238:c.113G>C p.A37V	rs2029810	0.01	0	0.000000	1	0.007778
15	86897921	C	G	ABCE1	NM_001238:c.113G>C p.A37V	rs2029810	0.002	0	0.000000	1	0.007778
16	77395165	T	G	AGAMT12	NM_139355:c.410T>G p.R136T	rs18808171	0.005	0	0.000000	1	0.007778
16	77461353	A	C	AGAMT12	NM_139355:c.775T>G p.F259V	rs75101616	0.002	0	0.000000	1	0.007778
16	88555536	S	C	ZNF409	NM_001124464:c.1115T>C p.G355H	rs152269922	0.002	0	0.000000	1	0.007778
17	5445256	A	C	NR3P1	NM_001033059:c.710T>G p.F234V	rs74546195	0.002	0	0.000000	1	0.007778
20	25952412	G	T	VDR1	NM_001252171:c.131A>G p.D43E	rs5115782	0.004	0	0.000000	1	0.007778

11/14

[illegible]

Figure 7

Gene	Position	ref	variant	rsID	unadjORs	AdjustedOR	conservation	RiskScores
PIKFYVE	209204738	T	G	rs61752191	7.445221988	10.99075775	1.061	65.94454648
GSN	124072992	G	A	rs41305623	7.297884642	10.92000131	1.048	65.52000786
ABCB5	20698170	T	G	rs199799182	3.998479589	7.098680693	1.043	56.78944555
COL5A1	137708884	C	T	rs61739195	4.523907939	8.320947224	0.927	49.92568334
CHST6	75513395	A	G	novel	6.060676854	7.640481078	0.833	45.84288647
PRDM5	121739667	T	A	rs146268537	6.509351654	11.36403959	1.061	45.45615835
ZEB1	31810782	A	C	rs118020901	4.938973916	6.70820032	1.088	40.24920192
MAP3K19	135744373	G	A	rs55867035	4.843600489	7.493452549	1.048	37.46726274
ABCB5	20793011	G	C	novel	2.082456433	4.91028734	1.048	34.37201138
AGBL1	86813160	G	A	rs199769049	2.082456433	5.583851032	1.048	33.50310619
ABCB5	20782629	A	G	rs189610663	1.711374705	4.068369126	1.088	32.54695301
COL8A2	36565674	C	T	rs779010935	2.392104822	4.468752836	0.807	31.28126985
PXDN	1652088	G	T	novel	2.56661194	4.431902898	1.048	31.02332029
CHST6	75512734	C	A	rs140699573	1.8720208	4.195130203	0.729	29.36591142
ZEB1	31809258	T	C	rs747652843	3.998479589	7.168892246	1.061	28.67556898
RAF1	12626632	T	G	rs74576012	1.794445831	4.086675957	1.061	28.6067317
KRT12	39022440	T	C	rs147584242	2.082456433	4.022893898	1.061	28.16025729
LTBP2	74968234	C	T	rs779067555	2.56661194	3.958356999	0.935	27.70849899
COL4A3	228172594	T	C	rs200302125	1.604740677	3.424322794	1.061	27.39458236
COL5A1	137688701	A	G	rs61736966	3.468364707	8.953770474	1.185	26.86131142
ABCB5	20793113	A	G	rs58795451	1.046941737	3.349730584	1.199	26.79784467
COL5A1	137706719	C	G	rs140797509	2.114088993	4.394124008	0.843	26.36474405
COL8A2	36565837	C	T	rs115156902	3.998479589	6.269469962	0.807	25.07787985
COL4A4	227973980	G	A	rs737557592	2.307095162	4.990955086	0.953	24.95477543
MTOR	11272469	C	T	rs768076455	4.064729967	4.981143262	0.852	24.90571631
TLN1	35719525	G	T	novel	1.683748741	4.146703411	1.048	24.88022046
GSN	124073007	G	A	rs368197143	1.82601854	4.823411974	1.048	24.11705987
MTOR	11189856	C	T	rs139043855	2.082456433	4.810711794	0.935	24.05355897
ABCB5	20784936	C	A	rs144219498	1.308920715	3.400600861	0.935	23.80420602
COL4A3	228158021	C	T	rs55816283	2.142152929	3.920587843	0.935	23.52352706
AGBL1	86800154	C	T	rs185896474	2.082360271	3.825850037	0.935	22.95510022
LTBP2	74968213	C	G	novel	2.082456433	4.385626379	0.935	21.9281319
ITGB1	33209254	A	C	rs201578528	1.865372055	5.454938693	1.14	21.81975477
VSX1	25058398	T	C	rs148957473	1.604740677	3.052325405	0.946	21.36627783
MYLK	123419283	G	C	rs200423083	3.998479589	7.071660106	1.036	21.21498032
TGFB1	135394841	G	A	rs200887459	1.154942389	2.946252613	1.048	20.62376829
KRT12	39023030	C	G	rs58038639	2.082456433	2.477953003	0.069	19.82362403
LTBP2	74976898	G	A	rs149991486	2.279315324	3.964228931	0.953	19.82114466
SHC1	154942911	G	A	rs115641580	1.864098582	3.952476297	0.953	19.76238149
ZNF469	88504062	A	G	novel	1.308920715	3.943040553	1.083	19.71520277
ABCB5	20685484	C	A	rs2074000	2.283378403	3.883383715	0.935	19.41691858
ZEB1	31810385	C	G	novel	1.711374705	3.21311931	0.852	19.27871586
KRT12	39021137	G	A	rs61735165	1.21045438	3.153619442	1.048	18.92171665
VSX1	25058389	G	C	rs576300014	1.254396383	2.697861991	0.935	18.88503393
ABCB5	20793107	C	A	rs377131299	1.787209265	3.138963926	0.935	18.83378356
COL5A1	137721882	G	C	novel	2.082456433	3.611060581	0.953	18.0553029
DIAPH1	140908492	C	A	rs370578225	1.473988805	3.000034183	0.935	18.0002051
IL1A	113539246	C	T	rs3783531	1.689042027	3.587871277	0.935	17.93935639
MTOR	11272422	G	A	rs746532612	1.49705933	2.911007097	0.953	17.46604258
ZNF469	88501431	C	A	rs201943633	3.998479589	4.346781561	-0.044	17.38712625
PTK2	141711018	C	T	rs200596356	1.49471852	2.874564966	0.935	17.2473898
COL5A1	137693848	C	T	novel	1.357747888	2.866115753	0.843	17.19669452
MTOR	11298591	G	A	rs780398729	1.041086412	2.863362877	1.048	17.18017726
ABCB5	20691167	G	A	rs151326874	1.111915537	2.777956166	0.993	16.66773699
ITGB1	33209192	T	C	rs372397463	1.346903435	2.761361763	1.011	16.56817058
COL4A4	227973559	G	T	rs749481247	1.308920715	3.290212458	0.902	16.45106229
ZNF469	88503039	A	C	rs139565761	1.308920715	3.962805126	1.088	15.8512205
ADAMTS18	77465400	G	A	novel	1.341477853	3.157090028	0.998	15.78545014
ROCK1	18619444	G	A	rs113763420	1.046941737	2.619228791	0.953	15.71537275
TLN1	35712890	T	A	rs756028633	1.467040886	3.142039112	1.061	15.71019556
LTBP2	74973503	T	G	rs761055280	1.473988805	3.088308972	0.964	15.44154486
PXDN	1651959	C	T	rs6723697	1.046941737	2.572504453	0.935	15.43502672
ITGB1	33218888	T	C	novel	1.308920715	3.669121058	1.011	14.67648423
KRT3	53186577	G	A	rs199614200	1.803075446	2.414136373	0	14.48481824
ZNF469	88502468	C	G	rs751901105	2.024731814	3.508898137	0.852	14.03559255
CYP4V2	187122332	G	A	rs34745240	16.40655143	13.89348218	0.953	13.89348218
MYLK	123385116	C	T	rs780094038	1.090838209	2.699693866	0.935	13.49846933
SHC1	154942864	G	A	rs563504056	1.117543639	2.691266377	0.953	13.45633189
ZNF469	88501344	G	A	rs547492890	1.782822912	3.360192143	0.902	13.44076857
PIKFYVE	209179036	G	A	rs146158348	1.154942389	2.679242385	0.953	13.39621192
MYLK	123471343	C	T	rs779602599	1.052918689	2.665950592	0.935	13.32975296
TLN1	35706215	T	C	novel	1.346903435	2.613587897	0.956	13.06793948
FYN	112041046	G	A	rs762290632	1.418211561	3.233273169	1.048	12.9309268
COL4A4	227912247	G	A	rs79143859	1.117543639	2.959474903	1.048	11.83789961
PIKFYVE	209168938	T	C	novel	1.346903435	2.902939808	1.061	11.61175923

Figure 7 cont.

COL5A1	137727030	G	A	rs143859495	1.573609109	2.880022245	0.953	11.52008898
GSN	124083642	C	T	rs116185403	1.227180183	2.763027329	0.75	11.05210932
PXDN	1664701	G	A	rs373022614	1.924925542	1.807466353	-0.017	10.84479812
COL5A1	137726950	C	T	rs2229817	1.49471852	2.645609165	0.852	10.58243666
LTBP2	74967651	C	T	rs143010135	1.111915537	2.62141819	0.935	10.48567276
MYLK	123512528	C	T	rs767453947	1.041086412	2.557414444	0.935	10.22965778
IL1A	113535653	C	A	rs1801715	1.467040886	2.549435566	0.852	10.19774226
PXN	120652725	C	T	rs111674215	1.328544696	2.536643925	0.831	10.1465757
ZNF469	88505454	C	T	novel	2.307095162	3.367008187	0.807	10.10102456
KRT3	53184043	C	T	rs367763143	2.082456433	3.21504366	0.772	9.64513098
FYN	112041022	G	C	rs376255649	1.368071308	3.173268878	1.048	9.519806634
COL4A3	228142188	T	C	novel	1.467040886	3.142039112	1.061	9.426117335
ZNF469	88504766	C	T	rs200668806	5.380113481	4.604499134	-0.031	9.208998269
PXN	120653019	C	T	rs192670913	1.025676306	2.301614601	0.831	9.206458404
ZEB1	31810514	A	G	rs138507868	1.177127237	3.031148131	1.088	9.093444393
WNT9B	44952657	C	G	novel	1.341477853	2.964329077	0.935	8.892987232
TLN1	35704761	C	T	rs771374255	1.23224986	2.93999823	0.847	8.819994689
ITGB1	33214913	A	T	novel	1.729798129	1.757236511	0.195	8.786182556
KRT12	39017974	T	C	rs150149178	2.082456433	2.192140991	0.113	8.768563962
ZNF469	88496148	T	G	rs753664726	1.473988805	2.896849256	0.9	8.690547767
COL4A4	227872865	G	A	rs200973262	2.082456433	1.703765732	0.137	8.518828659
FYN	111983038	G	C	rs28763975	3.92270113	2.633880723	0.154	7.90164217
ABCB5	20767954	G	A	rs17143304	2.355592989	1.535984426	0.076	7.679922129
MAP3K19	135744416	C	G	rs1112542	1.346903435	2.559274834	0.935	7.677824502
PIKFYVE	209138313	G	A	rs199642033	4.529884427	7.506924039	1.048	7.506924039
LCAT	67973896	C	T	rs755270091	2.082456433	3.735347816	0.922	7.470695632
SLC4A11	3208967	C	T	rs34224785	1.096019601	2.472050348	0.847	7.416151043
NLRP1	5424936	C	T	rs114529583	4.186377725	3.698122133	-0.768	7.396244266
SHC1	154938662	T	C	rs8191979	8.452547266	7.340202295	0.118	7.340202295
ABCB5	20666226	A	G	rs201244034	1.513413173	3.547948014	1.088	7.095896029
LTBP2	74972796	C	T	novel	1.346903435	2.355431529	0.852	7.066294588
WNT9B	44949945	A	G	rs118185468	5.235665827	7.043191906	1.083	7.043191906
TGFB1	135391462	A	G	rs188677757	1.254396383	3.512962168	1.199	7.025924335
ZNF469	88495372	G	A	novel	4.164888985	7.021802026	0.998	7.021802026
PXN	120651699	C	A	novel	1.007933509	2.30692467	0.837	6.92077401
WNT9A	228109286	C	T	rs552502207	1.177127237	2.247776418	0.789	6.743329254
V5X1	25062560	G	A	rs369865672	1.864098582	2.210697578	0.084	6.632092734
COL4A4	227924228	G	A	rs36121515	1.49471852	1.307730722	0.067	6.538653609
PXN	120653371	C	T	rs200855138	1.54404072	3.18728751	0.831	6.374575021
COL4A4	227927294	C	T	rs34236495	1.49471852	3.115200019	0.935	6.230400038
ZNF469	88498277	C	T	rs761094844	1.740280194	1.5349284	-0.007	6.1397136
ZNF469	88504203	G	C	rs199528724	1.309996605	2.958473106	0.949	5.916946212
MAP3K19	135744007	T	C	rs3905317	1.346903435	2.902939808	1.061	5.805879616
KERA	91449949	A	G	rs377616201	3.205470899	5.675950174	1.199	5.675950174
PIK3CG	106509082	G	A	rs144565710	1.03344886	2.597448039	0.953	5.194896079
ADAMTS18	77401554	C	T	rs200404040	1.346903435	2.559274834	0.935	5.118549668
RHOA	49398403	G	A	novel	1.729798129	2.508644107	0.551	5.017288215
ZNF469	88504204	G	C	rs56236932	5.504470701	4.890054144	0.014	4.890054144
ZNF469	88495361	C	T	rs202205643	2.082456433	2.430047393	0.049	4.860094785
AGBL1	87066168	G	A	rs144777747	2.954390882	2.382412923	0.135	4.764825847
COL8A2	36564213	C	G	rs199786966	1.144729549	2.359501305	0.834	4.719002609
MAPK3	30133311	C	T	rs200617125	1.009661703	2.344843255	0.852	4.689686509
ABCB5	20682892	C	T	rs141511940	2.082456433	2.30230492	-0.005	4.60460984
ZNF469	88496681	G	A	rs117995699	1.689042027	1.144032232	-0.208	4.576128929
LTBP2	75018994	G	A	rs137854861	3.998479589	4.517087187	0.063	4.517087187
COL5A1	137591755	C	T	rs41306397	3.205470899	2.254839914	0.02	4.509679827
LTBP2	75017841	G	A	rs147534770	1.346903435	1.110404041	0.1	4.441616166
COL4A4	227872887	C	T	rs77104306	1.117543639	1.1073966	0.065	4.4295864
MYLK	123453061	A	G	rs3796164	5.808400022	4.332955894	-0.276	4.332955894
ADAMTS18	77468556	C	G	rs751818903	1.54404072	1.374610262	-0.01	4.123830785
ROCK1	18564483	G	C	rs4562542	1.963895092	3.878559436	0.998	3.878559436
ABCB5	20767981	T	C	rs117749018	2.082456433	1.930627453	0.022	3.861254906
ZNF469	88495094	T	C	novel	1.238309099	1.267184001	0.133	3.801552002
AGBL1	86697745	A	G	novel	1.158986639	3.718537372	1.199	3.718537372
ZNF469	88504239	G	A	rs75288466	1.21045438	1.207498449	0.088	3.622495348
KRT3	53183951	C	CGCTGCCACCGCTGAAAC	rs570613061	3.58380065	3.183962839	0.207102714	3.406840238
ABCB5	20666206	G	T	rs369958537	2.082456433	1.650223545	-0.338	3.300447089
DIAPH1	140953564	GGAGGAGGAGGA	-	rs766218160	2.89520377	3.06823822	0.207102714	3.283014895
ZNF469	88504578	G	A	rs199610834	1.513413173	3.099898634	0.953	3.099898634
MYLK	123419062	T	C	rs75370906	1.090838209	3.062214612	1.061	3.062214612
ZNF469	88496786	GGCGGC	-	rs775423936	3.312136684	2.816841544	0.207102714	3.014020452
ABCB5	20784910	T	C	rs60197951	2.760947676	1.476473208	-0.215	2.952946415
ZNF469	88504350	C	G	rs183437633	1.308920715	0.977215713	-0.312	2.931647138
COL4A4	227917090	T	C	rs80243096	1.404954624	1.392540836	-0.121	2.785081673
V5X1	25062342	G	T	rs6050307	1.489353914	1.276285464	-0.071	2.552570928
AGBL1	87099508	G	A	rs73459659	5.369607585	2.440852996	-0.983	2.440852996

Figure 7 cont.

COL8A2	36563777 G	A	rs117860804	1.096019601	1.203276084	0.127	2.406552167
DIAPH1	140953259 G	A	rs200606811	1.117543639	2.36555308	0.824	2.36555308
COL4A3	228029515 C	T	rs139271412	1.117543639	1.10518402	0.063	2.21036804
LTBP2	74975983 C	T	rs761500203	1.099224002	1.101726802	0.025	2.203453603
COL4A3	228029513 C	G	rs184704920	1.82601854	1.101289269	-0.429	2.202578538
MTOR	11307762 T	C	rs146393749	1.534103874	2.116584728	0.144	2.116584728
IL6	22767134 C	A	rs142759801	1.090838209	0.699869341	-0.415	2.099608024
GSN	124074641 A	G	rs11550199	2.082456433	2.053268302	-0.119	2.053268302
AGBL1	86687049 C	T	rs186767486	1.117543639	0.657711992	-0.456	1.973135977
ZNF469	88502583 C	T	rs536725615	1.275893308	0.968259966	-0.455	1.936519933
ZNF469	88502234 G	A	novel	1.887718324	1.834672844	0.032	1.834672844
ZNF469	88502972 CTCCCGGGAACACC	-	rs281865162	2.082456433	1.712756795	0.207102714	1.83264977
ZNF469	88497404 C	G	novel	2.082456433	1.815630549	-0.242	1.815630549
LTBP2	75052587 G	A	rs149952751	1.23682719	1.701930387	0.429	1.701930387
NLRP1	5463226 A	-	novel	1.49471852	1.504402901	0.207102714	1.609711104
PAX6	31814881 A	C	rs763811795	1.513413173	1.470308329	0.207102714	1.573229912
FOXE3	47882574 G	C	rs281865461	1.473988805	0.769221498	-0.426	1.538442996
PAX6	31814918 G	A	rs767105283	1.49471852	1.38819461	0.207102714	1.485368232
LTBP2	75078581 G	C	novel	2.024731814	1.470058311	-0.018	1.470058311
MYLK	123419782 G	A	rs3732485	2.950545845	1.46681334	-0.275	1.46681334
MBNL1	152150657 CTGCAGCTG	-	rs532866577	1.238309099	1.364652518	0.207102714	1.460178194
ZNF469	88500137 G	A	rs144986357	1.308920715	1.407692923	0.053	1.407692923
ZNF469	88502573 G	A	rs202129382	2.082456433	0.673904049	-0.938	1.347808098
COL4A4	227973575 CCGGAGG	-	novel	1.046941737	1.242322528	0.207102714	1.329285105
LTBP2	74995228 C	T	rs145507464	2.082456433	1.321602553	-0.357	1.321602553
PIK3R1	67575499 T	A	novel	1.254396383	1.223186824	0.144	1.223186824
ADAMTS18	77468362 G	C	rs766063430	1.154942389	1.202657946	0.152	1.202657946
ZNF469	88504673 G	T	rs199760004	1.308920715	1.174628749	-0.128	1.174628749
PXN	120653416 C	T	rs770650321	1.23224986	1.168135607	-0.076	1.168135607
DIAPH1	140953681 C	T	rs182139018	1.573609109	1.068010976	-0.039	1.068010976
ZNF469	88502882 A	C	rs141776185	1.117543639	1.052335369	0.014	1.052335369
NLRP1	5461675 G	T	rs201368394	1.177127237	1.035559095	0.014	1.035559095
ADAMTS18	77401479 G	A	rs35701343	1.717201969	0.999699969	-0.206	0.999699969
MYLK	123444833 C	T	rs199988497	1.341477853	0.490981282	-0.863	0.981962564
ZNF469	88495487 G	A	rs184458982	1.21045438	0.488483699	-0.817	0.976967398
IL6	22771038 A	T	rs2069860	1.534103874	0.962524087	-0.644	0.962524087
LTBP2	75052569 G	A	rs143282840	2.082456433	0.946370334	-0.727	0.946370334
CYP4V2	187118720 G	A	rs34331648	1.117543639	0.822844413	-0.232	0.822844413
ZNF469	88494769 C	G	rs770483852	1.534103874	0.780987458	-0.853	0.780987458
ZNF469	88504853 G	A	rs3812951	1.052918689	0.749432286	-0.334	0.749432286
IL1RN	113890284 G	A	rs45507693	1.090838209	0.626966846	-0.525	0.626966846
NLRP1	5463093 C	T	rs140628502	1.683748741	0.625823848	-0.843	0.625823848
ZNF469	88498704 G	A	rs748403721	1.308920715	0.500549962	-0.981	0.500549962
AGBL1	87217666 C	T	rs185919705	1.158986639	0.165023137	-1.916	0.495069412
ZNF469	88498076 C	A	rs145158875	1.308920715	0.427822084	-1.138	0.427822084
ZNF469	88504205 G	C	rs532857190	1.803075446	0.404680106	-1.786	0.404680106
ZNF469	88501038 G	A	rs111828626	1.130110896	0.395391108	-1.182	0.395391108
ZNF469	88505284 C	A	rs141042464	1.308920715	0.110576481	-2.491	0.110576481