Analysis of "Junk DNA" Patents

July 2004

Prepared by:

Carol Nottenburg PhD JD*+

and Jade Sharples PhD+

* Cougar Patent Law, Seattle WA
carol.nottenburg@cougarlaw.com

+CAMBIA, Canberra ACT AUSTRALIA

current addresses:

CN: Cougar Patent Law, Seattle WA USA
JS: British High Commission, Canberra ACT Australia

THE CONTENTS OF THESE PAGES SHOULD BE USED FOR INFORMATION PURPOSES ONLY AND SHOULD NOT BE SUBSTITUTED FOR LEGAL ADVICE

Analysis of "Junk DNA" Patents:

"Intron Sequence Analysis Method for Detection of Adjacent and Remote Locus Alleles as Haplotypes"

1. Objective

The focus of the present study is an analysis of the so-called "junk DNA" patents, invented by Malcolm Simons and owned by GTG (Genetic Technologies Limited), and the controversy surrounding the patents. These patents, more appropriately called "uses of non-coding DNA for detecting alleles", have captured the attention of the scientific world and the public more so than many other patents, regardless of their scope. The impetus behind the attention seems to be the aggressive licensing strategy and practices of the patent owner GTG, an otherwise little-known Australian company, and amazement by scientists that patent offices could grant such patents.

Because of the widespread interest in these patents and the alarm expressed about the breadth of coverage, we have chosen to present not only an analysis of these patents but also discussion of the legal issues surrounding them. With this paper, we hope to raise the awareness and knowledge base of patent legalities in the scientific community.

2. State of the art in 1989

Less than 15 years ago, the state of genome research and nucleotide sequence-based diagnostics was relatively new. While the first discovery of a molecular, completely hereditary trait, the ABO blood group, occurred in the early 1900s, subsequent discovery and cataloguing of additional polymorphisms[1] in humans was slow and tedious, limited primarily by the analytical techniques. Until the mid-to-late 1970s essentially all polymorphisms were identified using biological reagents (e.g., antibodies) or electrophoresis. The tediousness and difficulties inherent in these

[1] The word "polymorphisms" means "many forms". When used in a biological sense, polymorphisms are genetically determined differences. Although polymorphisms could encompass virtually any detectable trait, the polymorphisms of most interest and usefulness are those at the molecular level. Such polymorphisms are detected by differences in nucleotide sequence.

techniques is reflected in the detection of only about 250 polymorphic proteins up until the end of the 1970s.

The explosion of new molecular technologies that began in the mid-1970s had a profound effect on identifying polymorphisms. Two techniques in particular markedly advanced the ability to identify polymorphisms: restriction enzyme digestion and DNA sequencing technologies. With the increasing commercial availability of restriction enzymes and the successful application of whole genome Southerns for more complex organisms, restriction fragment length polymorphisms (RFLPs) became a popular and preferred way to detect differences between individuals. Anyone who has tried their hand at these techniques can attest however to the difficulties and frustrations in obtaining clean and consistent results. Luckily, in the late 1970s, DNA sequencing was pioneered, which escalated the number of polymorphisms catalogued. With the further advent of polymerase chain reaction (PCR) in 1985, the information explosion magnified.

Using these techniques, different types of polymorphisms were characterized. Single base changes in coding regions of genomic DNA, especially those that resulted in amino acid changes, have long been sought after as a source of polymorphism, especially for diagnostic testing. Restriction enzyme mapping, and especially DNA sequencing of genomic DNA, allowed analysis of non-coding DNA regions as well, expanding the available pool of polymorphisms and facilitating the discovery of new types of polymorphisms, such as variable lengths of repeat DNA. For diagnostic testing however, the main reliance remained on detecting nucleotide differences in the coding sequence of the genes themselves. These changes that result in a different amino acid basis can be the basis for various diseases and syndromes. For example, a single amino acid change in each chromosomal copy of beta-haemoglobin results in sickle cell anaemia. So to test for carriers of one chromosomal copy, because the nucleotide change was not detectable as a restriction site polymorphism, it was necessary to determine the actual sequence of the individual's beta-haemoglobin genes or to perform hybridization studies with oligonucleotides, a technique that is notoriously difficult to detect single base differences. The advent of polymerase chain reaction (PCR) vastly improved the ease and reliability of assaying single nucleotide polymorphisms. Oftentimes however, testing of family members was still required for a conclusive diagnosis.

With the exception of rare restriction fragment length polymorphisms in linkage disequilibrium with some important disease-associated alleles[2], non-coding DNA was not generally regarded as useful for detecting polymorphic coding alleles. Apparently only one group found value in finding and using non-coding region polymorphisms to indicate the haplotype of a gene family.[3]

3. Summary of the "non-coding sequence" invention

As much as 97% of a human genome does not encode a protein. Researchers initially referred to this non-coding DNA as "junk" because no function could be found for it[4]. More recently, however, this wasteland has proved to be a repository for a variety of functions that are part of normal, and even critical, cellular processes. For example, non-coding DNA contains regulatory elements like enhancers, silencers of expression, and may function to promote exon shuffling in evolution. But what subsequently surprised many was that the non-coding DNA, even beyond these regulatory elements, can have conserved sequence among individuals.

While studying haplotypes of the human MHC (major histocompatibility complex), which encodes products critical for self-recognition and is extremely useful in myriad diagnostic tests, Dr Malcolm Simons discovered that non-coding DNA, both intronic and intergenic DNA, in the region of MHC genes, had sufficient non-random sequence variation to be informative in individuals for surrogate typing of MHC genes.[5] Moreover, the polymorphisms in the non-coding DNA were also informative of MHC

---

[2] see for examples, Kan and Dozy *Proc. Natl. Acad. Sci. USA 75*: 5631-5635 (1978); Little et al. *Nature 285*: 144-147 (1980); Wainwright et al. *Am J Hum Genet. 41*:944-947 (1987).

[3] Coincidentally, the first author on this paper is also the author of this patent analysis. Nottenburg, St John and Weissman, *J. Immunol. 139*: 1718-1726. The use of non-coding region polymorphisms to mark a chromosome was dictated by the limitations of genomic cloning of large DNA fragments.

[4] The phrase "junk DNA" is attributed to Dr Susumu Ohno, a very highly-regarded researcher at the City of Hope in Duarte, California. In 1972, in an attempt to explain the paradox that there was much more coding capacity in genomes than the number of genes, Dr Ohno proposed that much of the genome of more advanced eukaryotes was functionless. He called this DNA "garbage" or "junk" DNA. (see, Gregory, T.R. (2002). The C-value enigma. Ph.D. Thesis, Dept. of Zoology, University of Guelph, Guelph, Ontario, Canada. 894pp, especially Chapter 1 for a historical account. www.genomesize.com/rgregory/thesis/.

[5] Dr Simons has thoroughly reviewed his travels in arriving at the subject invention (www.junkDNApatents.com). Only a brief synopsis will be presented here.

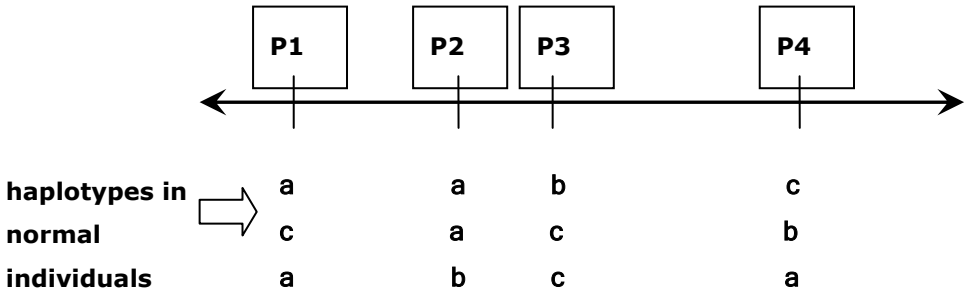
haplotypes[6]. From these discoveries, he recognized the universality of the discovery, namely, that the non-random, haplotypic structure of non-coding DNA would be a characteristic of the genomes of all eukaryotic organisms, rather than the alternative interpretation that the informative nature of HLA non-coding sequence was unique to HLA genes. This conclusion was influenced by previous RFLP (restriction fragment length polymorphism) studies of genes from other human loci, other animals, and of plants, in which restriction endonuclease cut sites had been shown to be linked to coding gene mutations/allelic characteristics.

Dr Simons also realized that various consequences flowed from these discoveries. To begin with, when combined with amplification, previously unknown non-coding polymorphisms in non-coding DNA could be captured. Moreover, the restricted heterogeneity in haplotypic structure of DNA in eukaryotic genomes would enable genome-wide mapping of gene regions associated with phenotypic traits, such as disease and drug responsiveness in humans and commercially desired characteristics in animals and plants. A big advantage of this technique is mapping by linkage disequilibrium (gametic association) in single individuals not known to be related by descent, thus avoiding the requirement for pedigrees.

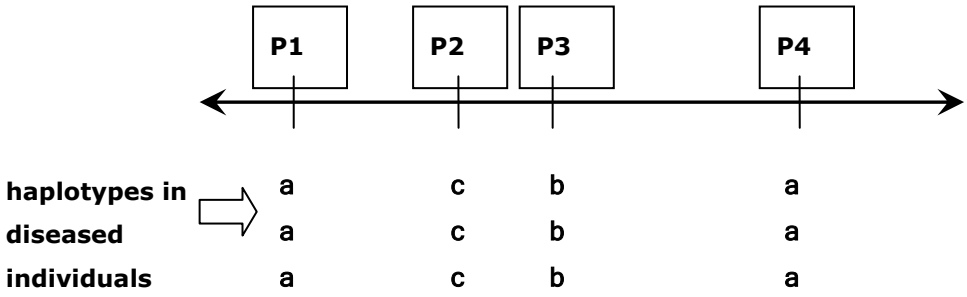
To illustrate the discovery, haplotype patterns of polymorphisms can be envisaged as in the exemplary drawing below. In this drawing of a chromosomal region, there are four polymorphic sites, P1 through P4. Each polymorphic site has three alleles, a, b, and c. If all possible haplotype patterns are found in a population, there would be 3^4 or 81 different patterns. In the drawing below, three of the possible patterns are shown.

---

[6] A "haplotype" is a set of closely linked genetic markers present on one chromosome which tend to be inherited together (not easily separable by recombination). The term "haplotype" has been extensively used in immunogenetics in referring to the linked genes in the MHC.
(see http://www.biochem.northwestern.edu/holmgren/Glossary/index.html).



Thus, to identify the chromosomal region containing the trait, haplotype patterns are determined for selected chromosomal regions. When a population is diverse, the haplotype patterns are expected to be diverse as well, such as above. When a population has a trait, then the region where the trait maps should show less diversity of haplotype patterns, such as shown below. Thus, the trait is mapped by comparing haplotype patterns for chromosomal regions between the diverse population and the population with the trait.



These realizations were captured in two patent application families initially filed in 1989 in the United States. Briefly, the two patent families are directed to (i) single gene allele and linked gene haplotype diagnostics, where the gene/region was known, and (ii) genome-wide gene discovery, where trait-associated genes had not been mapped. These patent families are discussed in more detail below.

4. Patent landscape overview

Some of the patents in the two patent families, owned by Genetic Technologies Ltd, are directed to intron sequence analysis methods for detection of haplotypes and genomic mapping. Other members of the families are directed to similar subject matter except

specifically claim the invention in relation to the human MHC (major histocompatibility complex) and, as such, will not be discussed herein.

The breadth of the claims in the patents listed below and their possible dominance over much of the current technology in molecular markers prompts a more detailed analysis of these patents. Furthermore, as mentioned above, one of the patents, US 5,612,179 (the '179 patent) is very contentious in the scientific research world.

The two patent families are complementary. The first family, exemplified by US 5,612,179, is directed toward detecting polymorphisms in non-coding regions of genomic DNA; the second family, exemplified by US 5,851,762 is directed to identifying polymorphisms that comprise a haplotype associated with a trait, such as a disease. As can be seen in the table below, these subject matters have been aggressively patented or pending in many countries, including Europe, Australia, Canada, Israel, Japan, and New Zealand.

Patent No.	Subject Matter	Comments
US 5,612,179	Methods for amplifying non-coding region DNA that encompasses an allele and detecting the allele.	Issued 18 Mar 1997 with 17 year term, but subject to terminal disclaimer; expires 9 Mar 2010. Related patents: AU 654111 B2, AU 672519 B2; IL 95467, JP 22417690; NZ 235051, SG 4700747; US 5192659; US 5789568; ZA 9006765; Related patents pending: CA 2023888, US 09/935,998
EP 0414469 B1		Registered in AT, BE, CH, DE, DK, ES, FR, GB, GR, HK, IT, LI, LX, NL, SE
US 5,851,762	Methods for mapping and identifying haplotypes of polymorphic markers in non-coding DNA, wherein the haplotypes are associated with a trait.	Issued 22 Dec 1998 with 17 year patent term; expires 22 Dec 2015. Related patents: AU 647086 B2; IL 98793; JP 3409796; NZ 238926; ZA 9905422 Related patents pending: CA 2087042
EP 0570371 B1		Registered in AT, BE, CH, DE, DK, FR, GB, IE, IT, LI, LX, NL, SE

5. Patent analysis: United States Patent No: 5,612,179

In this section, U.S. Patent No. 5,612,179 (the '179 patent) is dissected. A patent is composed of distinct sections: front page (bibliometric data), specification (text), and claims. The specification is further divided into the major sections: background, summary of the invention, description of the drawings, detailed description of the

invention, and examples.[7] A synopsis of the specification is presented first before turning to the claims and their interpretation.

The methodology of how to read and interpret claims is well-established. Claims are first read in light of the specification, but without drawing in any limitations set forth in the text. A thorough understanding of their meaning must include reading them in light of the prosecution history[8] as well. In contrast to the methodology, the rules that guide interpretation of the claims are constantly evolving. In the United States, the courts interpret the law and thus, essentially set the rules for interpreting claims. The specialized appeals court, the Federal Circuit, is the main court authority in the area. The system works similarly in Australia although the number of court cases are far fewer. In European countries, a centralized Patent Office prosecutes the patents but its appeals board lacks binding authority in the member countries. Thus, each member country could give a different interpretation of the same claims. For this report, interpretation follows the rules set forth by the European Patent Office Board of Appeals.

5.1. Bibliometric details

Title:	INTRON SEQUENCE ANALYSIS METHOD FOR DETECTION OF ADJACENT AND REMOTE LOCUS ALLELES AS HAPLOTYPES
Inventor:	Malcolm J Simons
Assignee:	Genetic Technologies, Ltd.[9]
Appl. No.:	08/949,652
Filed:	23 September 1992
Priority:	Earliest priority date is 25 August 1989
Issued:	18 Mar 1997
Expires:	9 Mar 2010[10]

[7] For information on the purpose of the various sections, see the tutorial "How to read a patent." on www.cambiaIP.org.

[8] The term "prosecution history" refers to the back and forth negotiation of the claim language and scope between the patent owner and her attorney and the Patent Office.

[9] The Assignee on the printed patent is GeneType AG, but due to corporate structure changes was re-assigned to Genetic Technologies Ltd., an Australian company.

5.2. Synopsis of the patent specification

The specification is directed to a method for detection of alleles (coding region polymorphisms) and haplotypes through analysis of polymorphisms of non-coding regions[11]. The method comprises amplifying genomic DNA with a primer pair, such that the amplified fragment contains at least some non-coding region DNA that is linked to an allele to be detected. Furthermore, the non-coding region has a polymorphism(s) that is assayed. In essence, detection of the polymorphism in the non-coding region is a surrogate for detection of the allele.

The text discusses details of the method, including (i) attributes of the amplification primers – the primers need to be specific to the genetic locus under investigation and preferably have an exact match to the genomic DNA sequence; (ii) the length of the sequence – in general, the more alleles there are, the longer the amplified sequence needs to be in order to distinguish among all alleles; (iii) the location of the amplified sequence –the non-coding region should be close enough to the allele to exclude recombination and loss of linkage disequilibrium between the amplified sequence and the coding region polymorphism; and (iv) detection methods – DNA sequence analysis, sequence specific oligonucleotide hybridization, and electrophoresis with or without restriction enzyme digestion. An additional part of the specification details the method with respect to MHC (major histocompatibility complex) genes. The examples demonstrate the invention in a forensic testing setting, paternity analysis, and for various MHC genes of human and cow.

5.3. Prosecution history

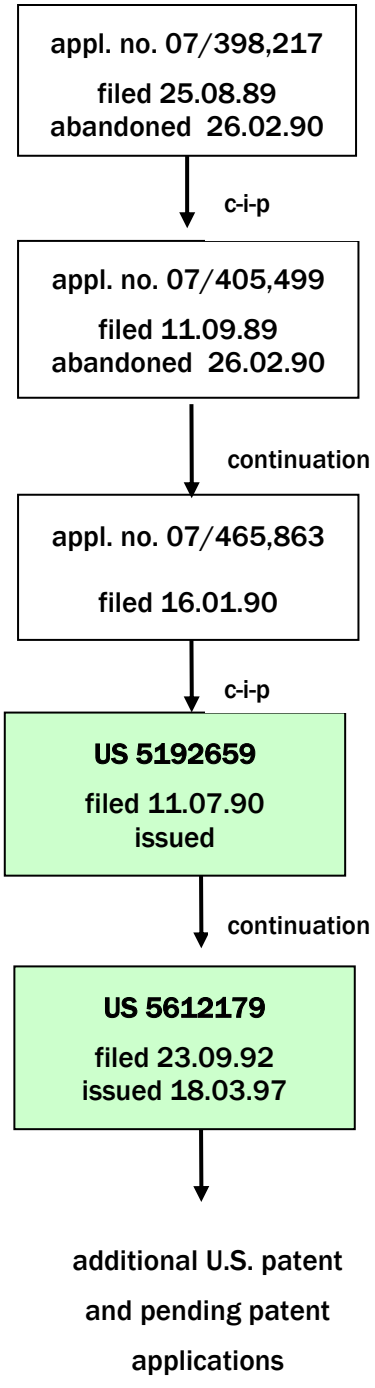
As seen in the diagram below, the '179 patent is a member of a family of patents and patent applications. The initial application was filed in August 1989. For help in

---

[10] This patent would have a term of 17 years from the date of issuance except that the term beyond 9 Mar 2010 was waived. Waivers (called terminal disclaimers) result from an "obviousness-type double patenting" rejection by the Patent Office. This means that at least one claim in this patent was deemed obvious over a claim in a related patent, which expires 9 Mar 2010. The rejection is overcome by disclaiming the length of term beyond the already granted patent. An additional requirement is that the two patents must remain co-owned.

[11] The patent text uses the term "intron" in an unconventional way – to mean non-coding DNA (*see* col. 5, lines 40-50).

interpreting the meaning and scope of the claims, prosecution histories are consulted from each of applications and patents in the boxes below.



5.3.1. US Application No: 07/398,217

The submitted claims were all directed to the invention as applied to the MHC locus. No substantive prosecution happened in this case. The application was abandoned in favour of the continuation-in-part.

5.3.2. US Application No: 07/405,499

The submitted claims were directed to the invention as applied to the MHC locus. No substantive prosecution happened in this case. The application was abandoned in favour of the continuation application.

5.3.3. US Application No: 07/465,863

The submitted claims were again directed to RFLP analysis of amplifying non-coding sequence of the MHC locus to identify MHC alleles. In a first examination, the Examiner believed that the claims encompassed two inventions[12]. The claims directed to methods and kits for MHC typing were chosen for prosecution and were rejected[13] as being obvious over a method for PCR-RFLP analysis of a c-ras gene (Deng and Mullis, *Nucl Acid Res 16*: 6231, 1988) in combination with either of two articles about a method for RFLP analysis of the MHC class II region.

The gist of the argument against this rejection is that (i) intron sequences are not random but "reflective of adjacent and remote allele sequences on the same chromosome"; (ii) amplifying intron DNA enables complete distinction between all alleles of a MHC gene locus; and (iii) the invention distinguishes heterozygotes and homozygotes.[14] The applicant then explains how prior to this invention, it was not believed that PCR and RFLP could be combined to distinguish MHC alleles. A lengthy exposition follows in which the applicant distinguishes his invention particularly from

---

[12] Only one invention can be granted protection in a patent. Thus, if the claims cover more than one invention, the applicant must choose one of the inventions to prosecute. The remaining inventions can be pursued in what are called divisional patent applications. The United States Patent and Trademark Office rules on multiple inventions are very strict. For example, nucleotides encoding a protein and the protein are considered two separate inventions.

[13] Only the main rejections of each examination report (Office Action) are discussed. Other rejections, if any, that are less germane to the thrust of this report are not discussed.

[14] Paper No. 7, Amendment, filed 12 Dec 1990, p. 6.

the cited prior art. The Examiner does not buy this argument however. Citing additional references demonstrating RFLP sites in intron sequences, the Examiner concludes basically that the Applicant is wrong.

In another response, the Applicant contends that "[u]ntil one recognizes that the intron sequences provide more variations and the variations correlate with the alleles, one cannot select a sufficiently informative sequence that is small enough to amplify."[15] This is the essence of the response to all subsequent art rejections.

After the Examiner repeats the rejection, the application is abandoned.

5.3.4. U.S. Application No: 07/551,239 (Patent No. 5,192,659)

Claims submitted in this continuation-in-part are directed to the more general case of using the invention to detect alleles of a genetic locus[16]. Claim 1 recited:

> A method for detection of at least one allele of a genetic locus comprising amplifying genomic DNA with an intron-spanning primer pair that defines a DNA sequence, said DNA sequence being in genetic linkage with said genetic locus and containing a sufficient number of intron sequence nucleotides to produce an amplified DNA sequence characteristic of said allele.

Without going into detail, the rejections for obviousness and the responses by t

he Applicant iterate those in the prior application. In response to citations of the prior art in which a "sequence polymorphism in an intron was empirically determined to correlate with a particular allele of a genetic locus, e.g. … β-thalassemia"[17], the Applicant amended the claim to the analysis of *multi-allelic* genetic loci. With respect to articles that describe RFLP sites in introns, the Applicant responded that because about 95% of DNA is non-coding, about 95% of restriction sites should be located there. But that fact did "not teach that intron sequences can be amplified and analysed to identify the alleles."[18] The Applicant stresses that it takes an exhaustive search to find an RFLP that correlates with a particular allele, in contrast to the invention in

---

[15] Paper No. 13, Amendment submitted 6 May 1991, p. 4.

[16] Claims are also directed to the specific case of the invention applied to MHC alleles.

[17] Paper No. 11, Amendment submitted 6 May 1991, p. 8.

[18] *Ibid,* p. 9.

which one only needs to know how the degree of polymorphism in order to amplify a sufficiently large piece of non-coding DNA to allow identification of all the alleles.

In the next Office Action, the Examiner additionally rejects the claims for lack of enablement, that is, the application does not teach how to use the invention for any region of DNA other than the MHC region. Despite protestations by the Applicant that the rejection should have been raised earlier and that the rejection lacks particularity, the Examiner refused to let up on either rejection. In response to the enablement rejection, the Applicant submits several declarations by prominent scientists who state that the generality of the invention was unexpected.

After several more rounds of rejections and responses, the claims were finally limited to the invention as it applied to MHC genes. The Examiner allowed these claims.

5.3.5. U.S. Application No: 07/949,652 (Patent No. 5,612,179)

The claims submitted in this application contains claims directed to the invention applied generally and claims directed to the invention applied to MHC genes. Claim 1 as submitted is the same as claim 1 of the patent except it lacks the analyzing step; claim 7 as submitted is the same as claim 9 of the patent.

Of interest, in a Preliminary Amendment[19] (an amendment submitted before the first Office Action), the invention is further characterized as not requiring information regarding family members. Furthermore, several methods specifically excluded from the claims are enumerated. These include "identifying a marker which could be used as a site of polymorphism to determine inheritance in family studies"; "sites of polymorphism in non-coding regions to attempt to track inheritance of disease genes in family studies"; classical RFLP analyses; and use of VNTR sequences as markers, such as for paternity testing and forensic applications.

A new Examiner in this case also rejected claims for lack of enablement and only rejected claims drawn to methods to determine RFLP fragments of an MHC locus for obviousness, although citing different prior art. While overcoming the lack of enablement rejection with the submission of declarations by prominent scientists, the

[19] Paper No. 5, Supplementary Preliminary Amendment, submitted 4 January 1993.

Applicant was unable to overcome the prior art rejection for obviousness. By canceling the rejected claims, the remaining claims should have been allowed.

The case was transferred to a new examiner however, who re-opened the prosecution and rejected the claims for lack of enablement. This remained a very difficult rejection to overcome as the Examiner criticized the statements in the Declarations and would not accept them as persuasive. Furthermore, the Examiner rejected claims as obvious over Kan et al. (identified RFLP near the β-globin gene) with a PCR reference. An interview with the Examiner and senior supervisor was used to discuss and resolve the issues of the Declarations and prior art. Following presentation of the interview as an amendment, the claims were allowed.

5.4. Claims at issue

There are eight independent claims in the '179 patent, but five of them (Claims 19, 24, 25, 33, and 36) are drawn to methods specific to the MHC locus of humans and will not be discussed. Independent claims 1 and 26 are analysed here as key claims; independent claim 9 is touched upon in this analysis where appropriate.

Claim 1 of the US '179 patent recites:

> A method for detection of at least one coding region allele of a multi-allelic genetic locus comprising:
>
> (a) amplifying genomic DNA with a primer pair that spans a non-coding region sequence, said primer pair defining a DNA sequence which is in genetic linkage with said genetic locus and contains a sufficient number of non-coding region sequence nucleotides to produce an amplified DNA sequence characteristic of said allele; and
>
> (b) analyzing the amplified DNA sequence to detect the allele.

Claim 9 is written in very nearly the same language as Claim 1. In the preamble (the initial statement prior to the word "comprising"), the method is recited as "for detection of at least one allele…", in step (a), the DNA sequence is in "genetic linkage with said allele", and in step (b), the amplified DNA is analysed to "determine the presence of a genetic variation in said amplified sequence to detect the allele." As discussed below, these appear to be differences without a distinction – meaning that the claims may have identical interpretations.

Claim 26 recites:

> A DNA analysis method for determining coding region alleles of a multi-allelic genetic locus comprising
>
> identifying sequence polymorphisms characteristic of the alleles, wherein said sequence polymorphisms characteristic of the alleles are present in a non-coding region sequence, said non-coding region sequence being not more than about two kilobases in length.

5.4.1. Meaning of the claims

In the United States, what the claims mean is a matter for the court (a judge) to decide[20]. In determining the meaning of claim terms, the focus is an objective test of what one of ordinary skill in the art at the time of the invention would have understood the term to mean. The written text of the patent may define the terms. Furthermore, the prosecution history (the back and forth negotiations between the patent applicant and the patent office) are considered as well. Like the specification, the prosecution history is used to ascertain or inform the meaning of the claim terms and is not used to enlarge or decrease the scope of the claims. Although a judge has the option to use other sources to help establish the true meaning of the claims, only the specification and prosecution are used in this report.

Before construing the claims, it may be helpful to the reader to review some special terms found in the claims whose meaning is commonly understood in patent law.

- "Comprising" means that all that follows is the minimum of what is claimed. Thus, if a method "comprises" steps A and B, a person performing the method using steps A, B, and C is infringing the claim.
- The indefinite article "a" or "an" means *one or more*.[21]

5.4.2. Claims 1 and 9

As an initial step to interpreting and understanding the meaning of claims, it is wise to pay close attention to the definitions within a patent. Conveniently, the drafter of the patent text put definitions of most of the key terms in a section entitled "Definitions", which can be found in columns 5 and 6 of the patent. Other definitions of terms are

---

[20] *Markman v. Westview Instr., Inc.,* 29 F.3d 1555 (Fed. Cir. 1994).

[21] Only in rare circumstances does the use of "a" or "an" limit the number to only one.

found in the prosecution history. For our purposes, only a few of these definitions are important. They are reproduced here and used below to help interpret the claims.

> "Allele" – a genetic variation associated with a coding region; … an alternative form of the gene. (In an Amendment[22], "allele" is further defined as "associated with a change in an exon sequence rather than a change in the sequence of the encoded protein or in non-coding regions of the gene sequence.")
>
> "Haplotype" – a region of genomic DNA on a chromosome which is bounded by recombination sites such that genetic loci within a haplotypic region are usually inherited as a unit.
>
> "Intron[23]" ("non-coding region sequence") – untranslated DNA sequences between exons, together with 5' and 3' untranslated regions associated with a genetic locus, as well as intergenic spacing sequences ("junk DNA").
>
> "Genetic locus" – the region of the genomic DNA that includes the gene that encodes a protein including any upstream or downstream transcribed non-coding regions and associated regulatory regions.
>
> "Genetic linkage" – regions of genomic DNA that are inherited together[24].

In plain English and in outline form, claims 1(and 9) can be written:

- A method for detection of at least one allele of a gene
  - the gene is multi-allelic (has two or more variant sequence)
- first amplify genomic DNA using a primer pair that amplifies non-coding DNA
  - the non-coding DNA is genetically and physically linked to the gene,
- then analyse the amplified DNA for a polymorphism
  - the polymorphism is indicative of a particular allele.

In other words, detection of the polymorphism in the amplified non-coding DNA acts as a surrogate for detecting the polymorphism in the coding DNA that is linked to the non-coding DNA.

---

[22] Paper No. 19, Amendment filed 12 Feb 1992, p. 9.

[23] This definition is a classic example of a patentee drafting his own definition, an allowable condition except when the definition is contrary to common usage. Because the Examiner correctly determined that "intron" is commonly used to refer to DNA sequences only between coding regions of a gene, the term was changed to "non-coding region sequence", which is consistent with common usage.

[24] Paper No. 11, Amendment submitted 6 May 1991 in 07/551,239, p.5.

Turning now to interpreting the claims, and claims 1 and 9 in particular, the plain language of the claim read in context of the definitions set forth in the specification and the prosecution history reveal the likely meaning and scope of the claim. To illustrate how the interpretation is arrived at, we look first at what each term means.

Preamble: According to the specification, an "allele" is a genetic variation *associated with* a coding region. While this definition does not require *per se* that the variation is in a coding region, and the specification[25] contemplates other locations, both the claim language and the prosecution history support the conclusion that the only location of the variation is in a coding region. In the preamble (the part of the claim before "comprising") of claim 1, the allele is specifically limited to a "coding region allele". Notwithstanding the axiom that preambles do not import limitations onto the claim, in this case the use of "allele" in step (a) refers back to the preamble and as such would likely be found to be limited in the same way. Furthermore, during prosecution of the patent, the Applicant made the bald statement that "allele" is "associated with a change in an exon sequence" and not with a change in the non-coding regions of the gene sequence. With such a statement, the Applicant affirms that his view of the invention entails detection of a coding region allele.

The term "multi-allelic genetic locus" can be parsed in the following way. By the definition in the text of the patent, a "genetic locus" includes all sequence found in a primary transcript (the exons and introns of a gene sequence and upstream and downstream transcribed non-coding regions), as well as regulatory regions. While promoter sequences initially spring to mind as a regulatory region, other sequences, which may be located at some distance from the transcribed region, are probably also part of the genetic locus as defined. The breadth of this term is not overly significant in the view of the author because the term allele has been limited to coding regions. "Multi-allelic" means two or more alleles, without counting the normal or wild-type sequence as an allele. This interpretation comes not from the specification, which does not define the term, but from the prosecution history. To "further distinguish" the invention from a disclosure journal article (Kan et al.) that taught a RFLP associated

---

[25] In the section "Screening Analysis for Genetic Disease", the inventors state that the method can "detect genetic diseases that are not associated with coding region variations but are found in regulatory or other untranslated regions of the genetic locus."

with an allele of β-globin[26] gene, the Applicant added the limitation "multi-allelic" and commented that the cited article only discussed a "bi-allele". Presumably the Applicant meant that there was only either the mutation that caused the disease or the normal gene. From this then, "multi-allelic" has to mean two or more variants in coding regions as compared to the normal sequence.

Step (a): The next phrase that can cause some difficulty in interpretation is a "primer pair that spans a non-coding region sequence". What the plain language of the claim doesn't say is where precisely are the sequences that the primer pair anneals to. The definitions in the specification do define the term "intron[27]-spanning primers" as a "primer pair that amplifies at least a portion of one intron." Furthermore, the primers can be located in (or more accurately – anneal to) conserved regions of the intron or in exon sequences that are "adjacent, upstream or downstream."[28] What the difference is between "adjacent" and "upstream or downstream" exons is an interesting question, but the answer is not likely to significantly impact, if at all, the ultimate interpretation of the claim. Suffice it to say that the primers can anneal to either exon or non-coding sequences[29].

Two further limitations on the DNA sequence defined by the primer pair are: (i) that it is in genetic linkage with the genetic locus; and (ii) that it contains a sufficient number of non-coding region nucleotides to produce an amplified sequence characteristic of the allele.

First the DNA sequence located between the primers must be in genetic linkage with the genetic locus. "Genetic linkage" was defined by the Applicant during prosecution as "regions of genomic DNA that are inherited together." In the case when the DNA sequence between the primers is found within the genetic locus, this limitation is

---

[26] *supra*

[27] While the term "non-coding region" is not explicitly defined, the term "intron" refers to the untranslated DNA sequences between exons, along with 5' and 3' regulatory and transcribed, but untranslated, regions. Furthermore, the term intron encompasses DNA located between genetic loci. The Examiner objected to the non-conventional use of the term "intron"; the Applicant substituted the term "non-coding region for "intron". For the sake of this report, we assume that the terms "non-coding" and "intron" are synonymous.

[28] Column 6, lines 10-15.

[29] Nothing excludes the possibility that one of the primers anneals to an exon sequence and the other of the pair anneals to a non-coding region.

nonsensical. In the cases when the DNA sequence between the primers is located outside the boundaries of the genetic locus, then it must be sufficiently close that the sequence and the locus are inherited together. In fact, the inventors prefer that the amplified DNA be in a non-coding region that is adjacent to an exon, and particularly next to a "variable exon", which is one that has allelic variants. If the amplified DNA sequence does not include non-coding sequence next to a variable exon, the sequence *must be sufficiently close to the variable exon to exclude recombination*[30]. Thus, the claims seem to require that the DNA sequence between the primers be associated, and very tightly linked, with a coding region variant (allele)

Second, the DNA sequence between the primers must contain a sufficient number of non-coding region nucleotides to produce an amplified sequence characteristic of the allele. What this basically means is that the characteristic that signifies an allele is found in the non-coding region portion of the amplified fragment. The difficulty with understanding the meaning of the limitation is the term "characteristic of". The Examiner also took issue with this term, contending that it was vague. In response, the Applicant recited a dictionary definition[31], which, in our opinion, did not illuminate much as to the meaning in the context of the claim. Further on, a bit more insight (or confusion) is provided when the Applicant states that sequences characteristic of an allele are "DNA sequences present in only one allele of a genetic locus." This statement appears to be nonsensical (because alleles are coding region variants according to the Applicant).

When the whole application is taken together with the prosecution history, the most likely interpretation is that the non-coding portion of the amplified sequence has at least one distinguishing nucleotide that is present when a particular coding region allele is present. Types of characteristic sequences include "change in the length of the sequence, gain or loss of a restriction site or substitution of a nucleotide.[32]" Relevant to this clause in the claims, the length of the non-coding region that is needed to

---

[30] Column 9, lines 19-29.

[31] "A trait, quality or property or a group of them distinguishing an individual, group, or type." Webster's Dictionary.

[32] Column 4, lines 44-48.

distinguish alleles is varies according to how many alleles there are. The more alleles, the longer the non-coding region that needs to be assessed.[33]

Finally, the claim requires "analysing the amplified DNA sequence to detect the allele." This is pretty much a throw-away statement, because without analysis to find the "characteristic sequence" that denotes a particular allele this claim has little value.

In plain terms, Claim 1 protects a method for detecting a coding region polymorphism or mutant (allele) of a genetic locus that has at least two alleles (other than the normal gene). The method comprises starting with genomic DNA and amplifying non-coding region DNA that is located close enough to the allele that the allele and the non-coding region DNA are co-inherited. Furthermore, the length of non-coding region DNA needs to be long enough to contain a sequence variation that correlates with a particular allele. Finally, when a sequence variation is found, then the presence of the correlated allele is proven.

The differences in claim 9 do not dramatically change the meaning from that of claim 1. It may be possible that claim 9 applies to bi-allelic loci, that is genetic loci that have a normal and a polymorphic or mutant allele. As discussed above, "genetic linkage with said allele" makes more sense than the phrase used in claim 1, but does not likely alter meaning. Finally, the "characteristic" is more or less specified in clause (b) as a "genetic variation", which is what the interpretation of "characteristic" was in claim 1.

Possibly more important than what the claim means is what types of non-coding region polymorphisms appear to be excluded from the claims. In the opinion of the author, such polymorphisms include those that are randomly situated (i.e., polymorphisms not linked to a particular gene), those that may map near a known gene but the gene does not exhibit allelic variants (e.g., a SNP that is located near or within an invariant gene), and those associated with a phenotype but not associated with a protein. Moreover, from the prosecution history (discussed above), some other methods that exploit non-coding region polymorphisms are explicitly excluded. These methods are those that rely on family data; "identifying a marker which could be used as a site of polymorphism to determine inheritance in family studies"; "sites of polymorphism in

---

[33] See section "Length of sequence" bridging columns 7 and 8.

non-coding regions to attempt to track inheritance of disease genes in family studies"; classical RFLP analyses; and use of VNTR sequences as markers, such as for paternity testing and forensic applications.

6. Summary

The so-called "junk DNA" patents are controversial, and the meaning of the claims is contentious. In this report, the claims of the United States patent are dissected and interpreted according to current U.S. patent law. As discussed, the patents disclose an invention that uses polymorphic sequences in non-coding regions to assess the presence of a polymorphism in an exon of a linked coding sequence. In contrast to the assertion that the patent claims all non-coding region polymorphism, the requirement in the claims for linkage of the two polymorphisms limits the scope of the patent.