Trilateral Project DR2
Report on Study on Single Nucleotide Polymorphism (SNP)/Haplotype Databases and Search Tools for Examiners

European Patent Office
Japan Patent Office
United States Patent and Trademark Office

I. Introduction

The promise of the development of designer therapeutics based upon genetic diversity has sparked great interest in patent protection on variations of genomic DNA sequences among individuals, especially single nucleotide polymorphisms (SNPs) and combinations of some SNPs known as haplotypes. Current technology allows laboratories to rapidly identify vast numbers of SNPs in short order. Expanding SNP and haplotype technology has resulted in an increase in the number of patent applications claiming SNPs and haplotypes, as well as corresponding methods of use. Furthermore, patent applicants often file applications that disclose and claim hundreds or thousands of related nucleic acid molecules.

II. Technical background

A comparative study of the challenges faced by the Trilateral Offices when examining claims directed to SNPs and haplotypes has been issued under Trilateral Project WM4 (Comparative studies in new technologies; Report on comparative study on examination Practice Relating to Single Nucleotide Polymorphisms (SNPs) and Haplotypes). Amongst topics raised, the report outlined that the prior art lacks any standardized naming, numbering, or characterization schemes for any particular gene or protein, especially when it is newly or recently discovered. This document also reported that differences in the manner in which the prior art and the application at issue describe/define a polymorphic site and/or a reference sequence make it difficult to perform a comprehensive search using textual databases or sequence databases. The search for a haplotype is even more complex than the search for a SNP because it is necessary to search for the presence of multiple polymorphic nucleotide positions within a single molecule. In addition, selection of appropriate databases is far from trivial, especially with respect to searching for an association between a haplotype and a patient's response to treatment by a drug.

III. Analyses Common to All Cases

Taking into account the difficulties encountered in searching claims directed to SNPs and haplotypes, the Offices agreed to exchange information to identify ways to improve searching efficiencies by exchanging methodologies for searching SNPs and identifying relevant databases.

For each of the PCT applications and claims set forth in section IV below, each Office was asked to provide the following:

Exchange/Communicate methodologies for searching single nucleotide polymorphisms (SNPs) and haplotype inventions.

1. Search claim(s) of specified published PCT application(s) to address the issue in the context of a specific example.
2. Include information relating to algorithms, databases, and search queries used, as well as the results set(s) obtained.

In addition, each Office was asked to identify SNP databases that are available or potentially available to the Office including access costs. Where available, information was requested relating to database preference, format, and completeness, as well as which databases were considered to be required to conduct a complete search, taking into account information redundancy between databases. The results of this portion of the study will appear in a separate report.

IV. Identification of PCT Applications and Claims

1. WO 03/051174 (26.06.2003) (PCT/US02/36095), Claims 52-56, wherein the polymorphism or mutation is at position 6i, 12e, 14e, or 14i.1 (as identified in Table 1 on page 12 of the description). Note that the nomenclature used in the claims is discussed in the description at page 10, lines 1-18, and Table 1, pages 12-13.

2. WO 01/79219 (25.10.2001) (PCT/US01/11853), Claim 21 wherein the haplotype is haplotype 1, 4, or 10, and Claims 10-12.

See Annex 1 for the claims, tables, and sequences, as well as interpretation information.

V. Results

Methodologies for searching single nucleotide polymorphisms (SNPs) and haplotype inventions.

1. Algorithms and Databases:

- ***USPTO***

a. In-house nucleic acid sequence databases (ABSS) using Smith-Waterman algorithm (GenBank, N-Geneseq, EST database, issued US patents and published US applications)

b. Chemical Abstracts Registry and bibliographic files

c. Text-based searches using USPTO's EAST tool (US patents and published US applications, EPO and JPO published applications/patents, Derwent)

d. Internet databases (Genecards, OnLine Mendelian Inheritance in Man (OMIM)

- ***JPO***

a. Chemical Abstracts Registry and bibliographic files

- ***EPO***

Sequence search algorithms employed FASTA (except for CAS searches, which used the Chemical Abstracts online proprietary search algorithm).

a) Nucleic acid sequence searches

- EMBL (EMBL contains all patent and non patent entries, for detail see: http://www.ebi.ac.uk/embl/Documentation/Release_notes/current/relnotes.html)
- Geneseq
- HGVBase (SNP database of EBI)
- CAS [Registry]

(b) Protein searches

- Protein databases provided by EBI (UniProt, aaGeneseq, Euro Patents, Japan Patents, US Patents, PDB).
- CAS [Registry]

(c) Sequence related search tools and databases (searched by keyword)

- Genecards (Weizmann Inst)
- LocusLink (NCBI): From LocusLink there is also a link to OMIN, which as well can comprise additional information in regard to SNPs/variants of the genes of interest.
- SRS (EMBL, Geneseq, HGVBase, RefSeq)

(d) Text based searches carried out in all searchable databases, either internally (EPOQUE, internal sessions) and via external hosts.

- Patent databases used: WPI, PAJ (Patent Abstracts of Japan), EPO internal databases.
- General databases: Biosis, Medline, Embase, Chemical Abstracts.

2. Search queries

The USPTO, JPO, and EPO used a combination of text (keyword) and sequence searches. The sequence searches done by the USPTO, JPO, and EPO included using fragments of each sequence that included SNP/mutation site as a search query, rather than the full-length sequences. In addition, the EPO used a search query directed to fragments of the proteins that included variant amino acids.

3. Results sets

Summary for WO 03/051174*:*

The USPTO, EPO and JPO identified the same reference (GenBank BF689594) for SNP 12e. The USPTO and EPO identified one additional reference for SNP 14e (GenBank BF880179). The EPO identified Saunders et al. Human Molecular Genetics 12:2765-2776 (2003) as applicable to SNPs 6i, 12e, and 14i.1 and WO 01/18250 as applicable to SNP 12e. The EPO also found some results with mutations within the claimed introns/exons, though not on the specific positions as given in Table 1 of WO 03/051174.

Observations*:* None of the USPTO, JPO, or EPO searches directed to specific SNPs identified WO 03/051174 as a relevant reference as there is no specific SEQ ID NO in the WO document that includes the mutant DNA; only the reference DNA sequence is present in the sequence listing.

Summary for WO 01/79219:

- The JPO's search identified two documents for Haplotype 2, WO 01/79219 (the application under consideration) and GenBank accession number AC084057.

- The USPTO identified four applicable references.
  WO 00/50436 (general disclosure of haplotyping method)
  GenBank F27586 (specific polymorphism at PS15)
  GenBank AA333722 (specific polymorphism at PS11)
  Bartel et al., (1993) Am J Hum Genet. 52(5):928-936 (specific polymorphism at PS11).

- The EPO identified the following references applicable to the indicated polymorphisms:

PS1

HGVbase/GeneCard-LocusLink: SNP002269848: dbSNPrsID record::rs3757869 and EMBL: EM_HUM: HSGNACHE/AC: L06484

PS10

HGVbase: SNP000000079

PS11

HGVbase/GeneCard-LocusLink: SNP000000080: dbSNPrsID record: rs7636; Ehrlich, Genomics 22:288-295,1994

EMBL: EM_HUM: HSACHEB/AC:L42812; Bartels et al., Am. J. Hum.Genet. 52(5):928-936 (1993)

GenBank AA333722/EMBL: EM_EST: HSZZ38773

PS15

EMBL: EM_EST: HSPD15598

REGISTRY: GenBank F27586 (EST, identical to EM_EST:HSPD15598)

Observations*:* Both the USPTO and the EPO identified the GenBank F27586, GenBank AA333722, and Bartels et al., references.

VI. Conclusions

Sequence searches for specific SNPs are not sufficient for a complete search because the variants are frequently not present in the full-length sequence, but rather are present in tables, text, or annotations. Furthermore, lack of standardized nomenclature and numbering systems make automated searching difficult as manual analysis and alignment of search results may be required. In addition, many databases are only text searchable, i.e., they do not permit direct sequence searches. A time intensive analysis may be necessary with respect to sequence variation information present in text or tables.

VII. Next Steps

To further the goal of improving searching efficiencies with respect to SNPs and haplotypes, there is a need for the Trilateral Offices to continue to evaluate available SNPs databases and to exchange information regarding the best databases for searching SNPs and haplotypes as new databases become available or existing ones are enhanced.

ANNEX 1 – Additional Information re PCT Applications and Claims

(1) – WO 03/051174 (26.06.2003) (PCT/US02/36095)

Claims

52. A purified or isolated nucleic acid comprising an alpha-2-macroglobulin sequence having a polymorphism or mutation at a position selected from the group consisting of 6i, 12i.1, 12i.2, 12e, 14e, 14i.1, 14i.2, 17i.1, 20e, 20i, 21i, 28i and 30e, wherein the nucleotide or nucleotide sequence at said position is other than an *A2M-1*.

53. The purified or isolated nucleic acid of Claim 52, wherein said alpha-2-macroglobulin sequence is SEQ ID NO: 1 or a sequence complementary thereto.

54. The purified or isolated nucleic acid of Claim 53, wherein the nucleotide or nucleotide sequence at said position is *A2M-2*.

55. The purified or isolated nucleic acid of Claim 52, wherein said alpha-2-macroglobulin sequence is selected from the group consisting of SEQ ID NOs: 2-8 and said polymorphism of mutation is at a position selected from the group consisting of 14e, 20e and 30e.

56. The purified or isolated nucleic acid of Claim 55, wherein the nucleotide or nucleotide sequence at said position is *A2M-2*.

Table

Table 1
Novel SNPs and Mutations Associated with Alzheimer's Disease

SNP/ Mutation	Location with reference to NCBI Accession Number AC007436 (SEQ ID NO: 1)	Location with reference to coding nucleotide sequences (e.g. cDNAs)	Nucleotide Change(s)	Amino Acid Change (with reference to SEQ ID NO: 9)
6i	174 bp downstream of exon 6 nucleotide position 37221		C A	
12e	exon 12 nucleotide position 45269	Nucleotide positions: 1339 of SEQ ID NOs: 3 and 5; and 1338 of SEQ ID NO: 7	C T	Y Y Silent effect
12i.1	152 bp upstream of exon 12 nucleotide position 45088		C G	
12i.2	115 bp upstream of exon 12 nucleotide position 45125		A T	
14e	exon 14 nucleotide position 47519	Nucleotide positions: 1730 of SEQ ID NOs: 3 and 5; and 1729 of SEQ ID NO: 7	T C	C R Amino acid position 563
14i.1	136 bp downstream of exon 14 nucleotide position 47669		insertion of AAG	
14i.2	151 bp downstream of exon 14 nucleotide position 47684		A C	
17i.1	240 bp upstream of exon 18 nucleotide position 53095		C G	
20e	exon 20 nucleotide position 56493	Nucleotide positions: 2574 of SEQ ID NOs: 3 and 5; 2573 of SEQ ID NO: 7; and 38 of SEQ ID NO: 4	C T	A V Amino acid position 844
20i	27 bp downstream of exon 20 nucleotide position 56586		C G	
21i	2 bp upstream of exon 21 nucleotide position 56887		T C	
28i	55 upstream of exon 29 nucleotide position 72076		G T	

Sequence

See the Figure of the application at:
http://l2.espacenet.com/espacenet/bnsviewer?CY=wo&LG=en&DB=EPD&PN=WO03051174&ID=WO++03051174A2+I+

Interpretation

For the purposes of this study, the following interpretation of claims 52-56 of WO 03/051174 applies:

Claim 52 is drawn to a purified or isolated nucleic acid that comprising any alpha-2-macroglobulin gene that includes a mutation in the cited positions.

By way of example, reference is made to the claimed polymorphism designated 6i.

> For claim 52 to the embodiment of a polymorphism/mutation designated 6i would encompass a purified or isolated alpha-2-macroglobulin gene in which the nucleotide at a position equivalent to the nucleotide 174 base pairs downstream of exon 6 nucleotide position 37221 of NCBI Accession Number AC007436 is altered to any nucleotide other than a C.
>
> For claim 53, the change is limited to SEQ ID NO: 1, with the exception that the position 174 base pairs downstream of exon 6 nucleotide position 37221 of NCBI Accession Number AC007436 is altered to any nucleotide other than a C.
>
> For claim 54, the change is limited to SEQ ID NO: 1, with the exception that the position 174 base pairs downstream of exon 6 nucleotide position 37221 of NCBI Accession Number AC007436 is altered to an A.

By way of example for claims 55 and 56, reference is made to the claimed polymorphism designated 14e.

> For claim 56, the claim is limited to a purified or isolated alpha-2-macroglobulin gene that includes one of SEQ ID NOs: 2-8 in which the nucleotide at a position equivalent to exon 14, nucleotide position 47519 of SEQ ID NO: 1 is altered to a C and is present within the context of one of SEQ ID NOs: 2-8.

(2) WO 01/79219 (25.10.2001)(PCT/US01/11853)
Claims

21. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the acetylcholinesterase (ACHE) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises an ACHE isogene defined by a haplotype selected from the group consisting of haplotypes 1-20 in Table 5; and

a second nucleotide sequence which is complementary to the first nucleotide sequence.

10. A method for predicting a haplotype pair for the acetylcholinesterase (ACHE) gene of an individual comprising:

(a) identifying an ACHE genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS11, PS12, PS13, PS14, PS15 and PS16;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the data in Table 4; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-16.

12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the acetylcholinesterase (ACHE) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-20 shown in Table 5 and the haplotype pair is selected from the haplotype pairs shown in Table 4, wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

Tables

Table 3. Polymorphic Sites Identified in the ACHE Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele	CDS Position	AA Variant
PS1	5197212	33850	A	C		
PS2	5197208	34137	C	T		
PS3	5197204	35944	C	T		
PS4	5197202	36030	C	T	36	S12S
PS5	5197200	36095	G	A	101	R34Q
PS6	5197198	36177	T	C	183	S61S
PS7	5197196	36832	A	G	838	T280A
PS8	5197194	36867	C	T	873	G291G
PS9	5197192	37026	G	A	1032	E344E
PS10[R]	5197190	37051	C	A	1057	H353N
PS11	5197184	37771	C	T	1431	P477P
PS12	5197178	38731	G	A		
PS13	5197176	40014	C	A		
PS14	5197174	40100	G	A		
PS15	8912014	40127	C	A		
PS16	8911825	40325	A	G		

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.
[R]Previously reported in literature.

Table 4 (Part 1). Genotypes and Haplotype Pairs Observed for ACHE Gene

Genotype Number	Polymorphic Sites										Hap Pair	
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10		
1	C	C	C	C	G	T	A	C	G	C	1	1
2	C/A	C	C	C	G	T	A	C	G	C	1	2
3	C	C/T	C	C	G	T	A	C	G	C	1	3
4	C	C	C	C	G	T	A	C	G	C	1	4
5	C/A	C	C	C	G	T	A	C	G	C	1	5
6	C/A	C	C	C	G	T	A	C	G/A	C	1	6
7	C	C	C	C	G	T	A	C	G	C/A	1	7
8	C/A	C	C	C	G	-	A	C	G	C	1	8
9	C	C	C	C	G	T/C	A	C	G/A	C	1	9
10	C	C/T	C	C	G/A	T	A	C	G	C	1	10
11	C/A	C/T	C	C	G	T	A	C	G	C	1	12
12	C	C	C	C	G	T	A	C	G	C/A	1	13
13	C/A	C	C	C	G	T	A	C	G	C/A	1	15
14	C	C	C	C	G	T	A	C	G	C	1	16
15	C/A	C	C/T	C	G	T	A	C	G	C	1	17
16	C/A	C	C	C/T	G	T	A	C	G	C	1	18
17	C/A	C	C	C	G	T	A/G	C	G	C	1	19
18	C	C/T	C	C	G	T	A	C/T	G	C	1	21
19	A	C	C	C	G	T	A	C	G	C	2	2
20	C/A	C/T	C	C	G	T	A	C	G	C	2	3
21	C/A	C	C	C	G	T	A	C	G	C	2	4
22	A	C	C	C	G	T	A	C	G	C	2	5
23	C/A	C	C	C	G	T	A	C	G	C/A	2	7
24	A	C	C	C	G	T	A	C	G	C	2	8
25	A	G	C	C	G	T	A	C	G	C	2	14
26	C	T	C	C	G	T	A	C	G	C	3	3
27	C	C/T	C	C	G	T	A	C	G	C	3	4
28	C	C	C	C	G	T	A	C	G	C	4	4
29	C/A	C	C	C	G	T	A	C	G	C	4	5
30	C/A	C	C	C	G	T	A	C	G	C	5	11

Table 4 (Part 2). Genotypes and Haplotype Pairs Observed for ACHE Gene

Genotype Number	Polymorphic Sites						Hap Pair	
	PS11	PS12	PS13	PS14	PS15	PS16		
1	C	G	C	G	C	A	1	1
2	C	G	C	G	C	A	1	2
3	C	G	C	G	C	A	1	3
4	C/T	G	C	G	C/A	A	1	4
5	C	G	C	G	C	A/G	1	5
6	C	G	C	G	C	A	1	6
7	C/T	G	C	G	C/A	A	1	7
8	C	G/A	C	G	C	A/G	1	8
9	C	G	C	G	C	A	1	9
10	C	G	C	G/A	C	A	1	10
11	C	G	C	G	C/A	A	1	12
12	C/T	G	C	G	C	A	1	13
13	C/T	G	C	G	C/A	A/G	1	15
14	C	G	C	G	C	A/G	1	16
15	C	G	C	G	C	A	1	17
16	C	G	C	G	C	A	1	18
17	C	G	C	G	C	A	1	19
18	C/T	-	C	G	C/A	A	1	21
19	C	G	C	G	C	A	2	2
20	C	G	C	G	C	A	2	3
21	C/T	G	C	G	C/A	A	2	4
22	C	G	C	G	C	A/G	2	5
23	C/T	G	C	G	C/A	A	2	7
24	C	G/A	C	G	C	A/G	2	8
25	C	G	C/A	G	C	A	2	14
26	C	G	C	G	C	A	3	3
27	C/T	G	C	G	C/A	A	3	4
28	T	G	C	G	A	A	4	4
29	C/T	G	C	G	C/A	A/G	4	5
30	C/T	G/A	C	G	C	G	5	11

Table 5. Haplotypes Identified in the ACHE Gene

Haplotype Number	Polymorphic Sites															
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16
1	C	C	C	C	G	T	A	C	G	C	C	G	C	G	C	A
2	A	C	C	C	G	T	A	C	G	C	C	G	C	G	C	A
3	C	T	C	C	G	T	A	C	G	C	C	G	C	G	C	A
4	C	C	C	C	G	T	A	C	G	C	T	G	C	G	A	A
5	A	C	C	C	G	T	A	C	G	C	C	G	C	G	C	G
6	A	C	C	C	G	T	A	C	A	C	C	G	C	G	C	A
7	C	C	C	C	G	T	A	C	G	A	T	G	C	G	A	A
8	A	C	C	C	G	T	A	C	G	C	C	A	C	G	C	G
9	C	C	C	C	G	C	A	C	A	C	C	G	C	G	C	A
10	C	T	C	C	A	T	A	C	G	C	C	G	C	A	C	A
11	C	C	C	C	G	T	A	C	G	C	T	A	C	G	C	G
12	A	T	C	C	G	T	A	C	G	C	C	G	C	G	A	A
13	C	C	C	C	G	T	A	C	G	A	T	G	C	G	C	A
14	A	C	C	C	G	T	A	C	G	C	C	G	A	G	C	A
15	A	C	C	C	G	T	A	C	G	A	T	G	C	G	A	G
16	C	C	C	C	G	T	A	C	G	C	C	G	C	G	C	G
17	A	C	T	C	G	T	A	C	G	C	C	G	C	G	C	A
18	A	C	C	T	G	T	A	C	G	C	C	G	C	G	C	A
19	A	C	C	C	G	T	G	C	G	C	C	G	C	G	C	A
20	C	T	C	C	G	T	A	T	G	C	T	G	C	G	A	A

Sequence

See Figures 1 (all) and 2 of the application at:
http://l2.espacenet.com/espacenet/bnsviewer?CY=wo&LG=en&DB=EPD&PN=WO0179219&ID=WO+□
++0179219A2+I+□