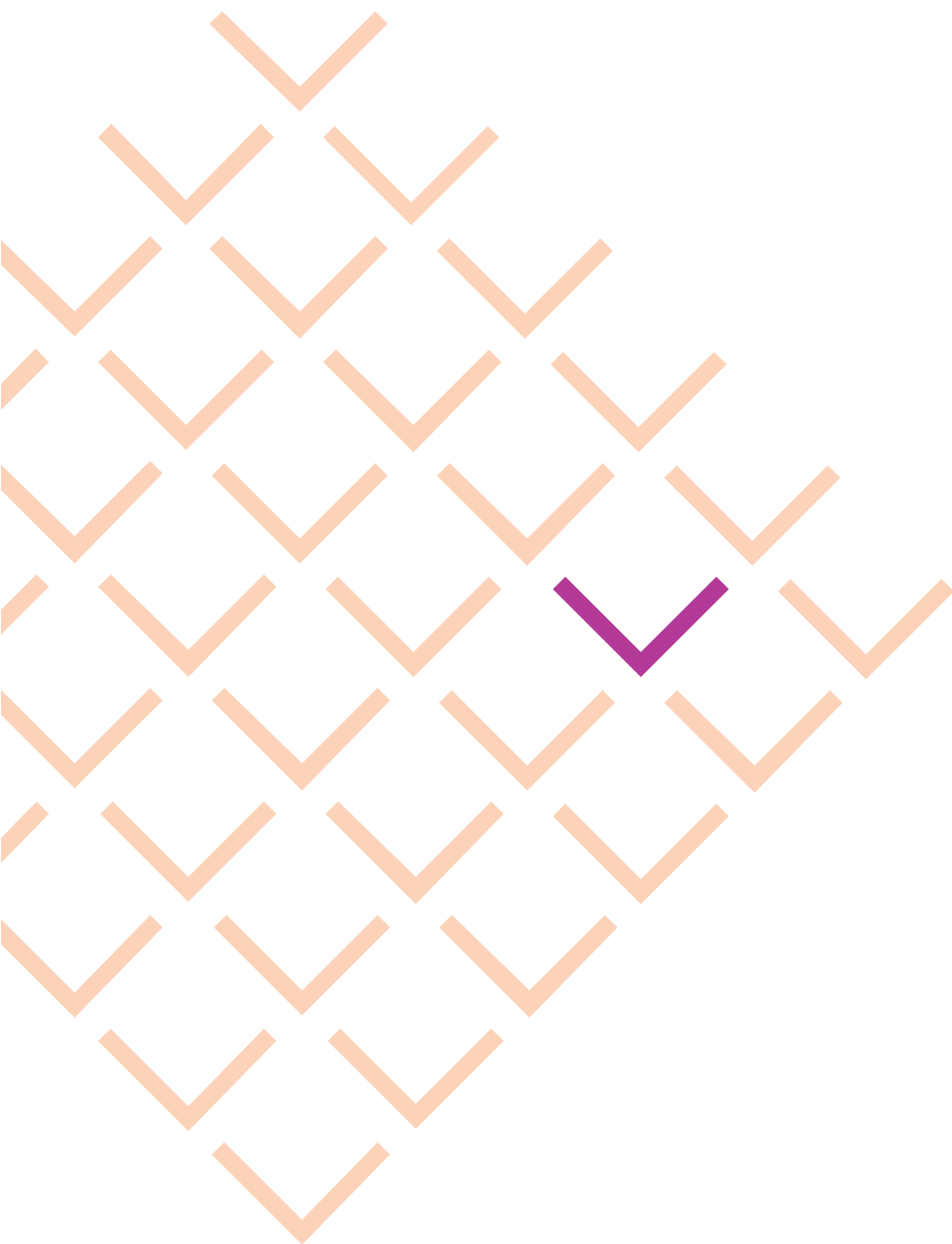




Nucleic Acid Sequences in the CAS Registry File on STN- Exact and Pattern Searching

A Quick Reference Guide



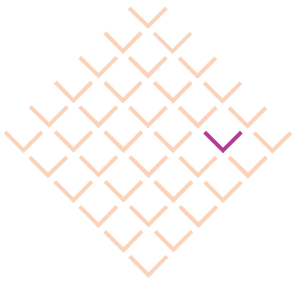
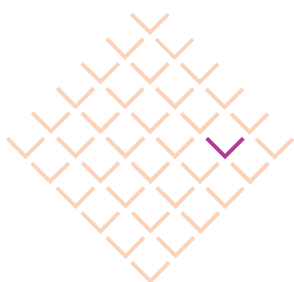


Table of Contents

Preface	3
Searching exact sequences	4
Searching partial sequences	6
Pattern searching	7



Preface

This booklet is designed as a quick reference guide to exact and pattern searching of nucleic acid sequences in the CAS Registry File on STN.

For more detailed information on searching in the Registry File on STN, please refer to the user manual *REGISTRY File: Biosequence Searching* available from Chemical Abstracts Service (CAS).

In addition, the database summary sheet for the Registry File provides examples of all the search and display options in this file. The summary sheet is available in print, online in STNGUIDE, and at: www.cas.org/ONLINE/DBSS/registryss.html.

You may also use extensive online help within each file on STN or contact the Help Desk.

For web access to quick reference guides, visit: www.cas.org/ONLINE/QRGUIDES/qrguide.html.

Help Desk

Phone: 800-848-6533 or 614-447-3698

Fax: 614-447-3798

E-mail: help@cas.org

Searching exact sequences

To search exact sequences of nucleotides, enter the sequence in the /SQEN field of the Registry file.

The following codes may be used in exact nucleic acid sequence searches:

Code	Name or Definition
A	adenine
C	cytosine
G	guanine
T	thymine (in DNA)
U	uracil (in RNA)
I	inosine

Find literature or patents on a diagnostic probe with the sequence CGCCCCTGCGTTACCCTCCCCGCCG

```
=> S CGCCCCTGCGTTACCCTCCCCGCCG/SQEN
L1      3 CGCCCCTGCGTTACCCTCCCCGCCG/SQEN

=> D SEQ NTE LC 3

L1 ANSWER 3 OF 3 REGISTRY COPYRIGHT 2003 ACS

SEQ      1 cgcccctgcg ttaccctccc cgccg
          =====
HITS AT: 1-25

**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
LC STN Files: CA, CAPLUS, TOXCENTER, USPATFULL

=> SEQLINK
ENTER TYPE OF LINK (EXACT) OR ?:EXACT
ENTER (L2), L# OR ?:L1
L2      3 SEQLINK EXACT L1
```

Enter REGISTRY.

Use the SEARCH (or S) command to search the exact sequence in the /SQEN field.

Display the sequence (SEQ), annotation (NTE), and the Locator (LC) field listing the files containing references to the CAS Registry Number.

Use the SEQLINK command (free of charge) to find related sequences, if any.

=> **FILE CAPLUS**

=> **S L2**

L3 1 L2

=> **D BIB AB HITSEQ**

L3 ANSWER 1 OF 1 CAPLUS COPYRIGHT 2003 ACS

AN 1995:884205 CAPLUS Full-text

DN 123:278057

TI Early diagnosis of breast cancer by analysis of patterns of gene expression and treatment using the BRCA1 gene

IN Holt, Jeffrey T.; Jensen, Roy A.; Page, David L.; Obermiller, Patrice S.; Robinson-Benion, Cheryl L.; Thompson, Marilyn E.

PA Vanderbilt University, USA

SO PCT Int. Appl., 97 pp.

CODEN: PIXXD2

DT Patent

LA English

FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 9519369	A1	19950720	WO 1995-US608	19950117
			:		
			:		
	US 6342483	B1	20020129	US 1998-7678	19980115
PRAI	US 1994-182961	A	19940114		
	US 1995-373799	A	19950117		
	WO 1995-US608	W	19950117		

AB A method of detecting and diagnosing pre-invasive breast cancer by identifying differentially expressed genes in early, pre-invasive breast cancer tissue is described. Differentially expressed genes

:

control tissue RNA. The present invention also provides a method of screening for compds. that induce expression of the BRCA1 gene, whose product neg. regulates cell growth in both normal and malignant mammary epithelial cells. The use of the BRCA1 gene in gene therapy is also discussed.

IT **169596-15-0**

RL: PRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses)

(PCR primer, in differential display diagnosis of breast cancer; early diagnosis of breast cancer by anal. of patterns of gene expression and treatment using BRCA1 gene)

RN 169596-15-0 CAPLUS

CN DNA, d(C-G-C-C-C-C-T-G-C-G-T-T-A-C-C-C-T-C-C-C-C-G-C-C-G)
(9C1) (CA INDEX NAME)

NTE singlestranded

SEQ 1 cgcccctgcg ttaccctccc cgccg

Enter one or more of the files containing the CAS Registry Number.

Search the REGISTRY L-number (L1).

Display the bibliographic information (BIB), the abstract (AB), and the index entry for the hit sequence (HITSEQ).

Searching partial sequences

To search partial sequences or sequences with gaps, repeating units, or alternate units, enter the partial sequence in the /SQSN (subsequence) field. You can use the codes for specific nucleotides or ambiguity codes.

Specific Code	Name or Definition
A	adenine
C	cytosine
G	guanine
T	thymine (in DNA)
U	uracil (in RNA)
I	inosine

Ambiguity Codes	Name or Definition
Y	pyrimidine (T or C)
R	purine (G or A)
M	amino (A or C)
K	keto (G or T)
S	strong interaction (3 H bonds) (G or C)
W	weak interaction (2 H bonds) (A or T)
B	not-A (G or T or C)
V	not-T, not-U (G or C or A)
D	not-C (G or A or T)
H	not-G (A or C or T)
N	unknown nucleotide (G or A or T or C)
X	uncommon nucleotide
Z	non-specific nucleotide

Pattern searching

Complex pattern searching of nucleic acid subsequences is possible using special notation for gaps, repeating units, and other types of variability.

Gaps

Symbol	Function	Example
.	a gap of one unit	TACGGGG.TG/SQSN
{m}	a gap of m units	CTCGTGATTA.{5}GG/SQSN
{m,u}	a gap of m to u units	ATGGC.{1,50}ATGGC/SQSN
.?	a gap of zero or one unit	GATTA.?TTG/SQSN
.*	a gap of zero or more units	ATCTTCCTGT.*CCCTC/SQSN
.+	a gap of one or more units	TACGG.+GAGAGCTT/SQSN

Repetition

Symbol	Function	Example
{ }	repeat the preceding unit with a number or range	GAAT(TAA){2}/SQSN
?	repeat the preceding unit zero or one time	CAT(CGA)?GGAC/SQSN
*	repeat the preceding unit zero or more times	CAT(CTG)*TATT/SQSN
+	repeat the preceding unit one or more times	CAT(CTG)+TATT/SQSN

Other variability options

Symbol	Function	Example
^	require the unit at the beginning or the end of the sequence	^GGAAGGG/SQSN CCTC^/SQSN
[]	specify alternate units	CATCTG[CG]C/SQSN
[-]	exclude a unit	TTGGG[-G]TTT/SQSN
	specify alternate sequences	TTA TTG/SQSN
&	join together sequence queries	L1&L2/SQSN (L1 and L2 are sequence queries)

Order of execution

More than one symbol may be used to create complex sequence queries. If you do not use parentheses in sequence queries, the operations will be executed in the following order:

1. repeat symbols ? or * or +
2. repeat expressions using curly braces, e.g., {3,6}
3. concatenation symbol &
4. the vertical bar |

Pattern searching

Find patents and literature on the following partial sequence: AGGGTATAAAAA...(CCA|ATG), where is a gap of 4 nucleotides followed by either CCA or ATG

=> **FILE REGISTRY**

=> **S AGGGTATAAAAA...(CCA|ATG)/SQSN**

L2 332 AGGGTATAAAAA...(CCA|ATG)/SQSN

=> **D KWIC SQL 1**

L2 ANSWER 1 OF 332 REGISTRY COPYRIGHT 2003 ACS

SEQ 51 cgggatggcc agggataaaa aagggccac aagagaccgg ctctaggatc
=====

HITS AT: 61-79

SQL 1008

=> **FILE USPATFULL CAPLUS BIOSIS GENBANK**

=> **S L2**

L3 12 FILE USPATFULL

L4 60 FILE CAPLUS

L5 8 FILE BIOSIS

L6 286 FILE GENBANK

TOTAL FOR ALL FILES

L7 366 L2

=> **SET DUPORDER FILE**

SET COMMAND COMPLETED

=> **DUP REM L7**

DUPLICATE IS NOT AVAILABLE IN 'GENBANK'.

ANSWERS FROM THESE FILES WILL BE CONSIDERED UNIQUE
PROCESSING COMPLETED FOR L7

L8 356 DUP REM L7 (10 DUPLICATES REMOVED)

ANSWERS '1-12' FROM FILE USPATFULL

ANSWERS '13-64' FROM FILE CAPLUS

ANSWERS '65-70' FROM FILE BIOSIS

ANSWERS '71-356' FROM FILE GENBANK

Enter REGISTRY.

Search the partial sequence in the /SQSN field.

Display the hit sequence fragment in context (KWIC) and the sequence length (SQL).

Enter the reference files with CAS Registry Numbers for nucleic acid sequences.

Search the REGISTRY L-number (L1). Each file is searched and an L-number answer set is created in each file. A composite L-number (L6) with all references is created.

Set the arrangement of answers in file order in the process of duplicate identification or elimination.

Remove duplicates.

Answers are arranged in file order.

=> **D TI PA AB HITRN**

L8 ANSWER 1 OF 356 USPATFULL DUPLICATE 2
TI Galanin transgenic mice
PA The University of Manitoba, Winnipeg, CANADA (non-U.S. corporation)
AB A transgenic mammal whose somatic and germ cells having a nucleic acid construct wherein the construct includes a mammalian promoter operably linked to a cDNA genomic sequence is provided for the overexpression of galanin. Also provided is a construct having cDNA for the overexpression of galanin. A method of making a transgenic mammal by producing a mammal having a construct for the overexpression of galanin is provided.
IT 439766-07-1
(nucleotide sequence; galanin transgenic mice as pituitary adenomas models)

=> **D L8 BIB AB 13**

L8 ANSWER 13 OF 356 CAPLUS COPYRIGHT 2003 ACS
DUPLICATE 1
AN 2003:270223 CAPLUS Full-text
DN 138:266964
TI Gene expression profiles useful in methods of diagnosis of cancer compositions and methods of screening for modulators of cancer
IN Afar, Daniel; Aziz, Natasha; Gish, Kurt C.; Hevezi, Peter A.; Mack, David H.; Wilson, Keith E.; Zlotnik, Albert
PA EOS Biotechnology, Inc., USA
SO PCT Int. Appl., 767 pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 21

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI WO 2003025138	A2	20030327	WO 2002-XD29560	20020917
	:			
	:			

AB Described herein are genes whose expression are up-regulated or down-regulated in specific cancers, including acute lymphocytic leukemia, glioblastoma, glioblastoma multiforme, glioma, kidney cancer, stomach cancer, melanoma, and benign NEVI. Mol. profiles of various normal and cancerous tissues were detd. and analyzed using the Affymetrix/Eos Hu01 and Hu03 GeneChip microarrays contg. 35,403 and 59,680 probe sets, resp. Related methods and compns. that can be used for diagnosis and treatment of those cancers are disclosed. Also described herein are methods that can be used to identify modulators of selected cancers. [This abstr. record is one of nine records for this documents necessitated by the large no. of index entries required to fully index the document and publication system constraints.].

Display references from each file or selected files.

Answer 1 is from USPATFULL.

Answer 13 is from CAPLUS.

=> **D 65 ALL**

Answer 65 is from BIOSIS.

L8 ANSWER 65 OF 366 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.

AN 1992:523573 BIOSIS Full-text

DN BA94:131648

TI CHARACTERIZATION OF AN ALTERNATIVE PROMOTER IN THE HUMAN GROWTH HORMONE GENE.

AU COURTOIS S J; LAFONTAINE D A; ROUSSEAU G G

CS UCL-ICP BOX 7529, 75 AVE. HIPPOCRATE, B-1200 BRUSSELS, BELGIUM.

SO J BIOL CHEM, (1992) 267 (27), 19736-19743.
CODEN: JBCHA3. ISSN: 0021-9258.

FS BA; OLD

LA English

AB Transcription of the human growth hormone (hGH) gene depends on cis-acting elements contained within 300 base pairs of its 5'-flanking sequence. An earlier in vitro study of the transcriptional activity of this 5'-flanking region suggested that transcription can start upstream from position +1. We have investigated this phenomenon by cell-free transcription and transient transfection of chimeric constructs in cultured pituitary cells and in HeLa cells and by analysis of RNA from human pituitary glands and HeLa cells. Transcription initiation sites were identified at positions -54 and -197 by cell-free transcription assays and by RNase mapping of human pituitary RNA. In transfection assays, the hGH gene 5'-flanking sequence upstream from position -197 displayed transcriptional activity, which critically depended on the upstream stimulatory factor-binding site located between positions -253 and -266. Transcripts initiated upstream from position +1 were detected in human pituitary RNA by polymerase chain reaction amplification and Northern blotting. These transcripts were longer than the mRNA encoding hGH. They might control initiation at position +1 or code for a novel peptide.

CC Genetics and Cytogenetics - Human *03508

Biochemical Studies - Nucleic Acids, Purines and Pyrimidines

10062 Replication, Transcription, Translation *10300

Endocrine System - Pituitary *17014

BC Hominidae 86215

IT Miscellaneous Descriptors

TRANSCRIPTION PITUITARY RNA MESSENGER RNA

MOLECULAR SEQUENCE DATA NUCLEOTIDE SEQUENCE

GENBANK-J03071 EMBL-J03071

RN 141145-46-2 (GENBANK-J03071)

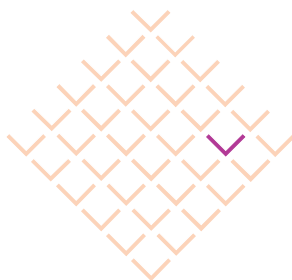
Help Desk

Phone: 800-848-6533 or 614-447-3698

Fax: 614-447-3798

E-mail: help@cas.org

Hours: 8 AM - 8 PM Eastern Time



Customer Service

Phone: 800-753-4CAS (4227)

614-447-3731

Fax: 614-447-3751

E-mail: help@cas.org



*CAS is a division of the
American Chemical Society.*

CAS1855-0503

May 2003