



(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/12* (2006.01)

S. Road, Taipei City, 106 (TW). **YU, Chong-Jen**; 2F, No. 8, Alley 10, Lane 244, Sec. 3 Roosevelt Rd., Taipei (TW).

(21) International Application Number:
PCT/US2010/047778

(74) Agent: **TSAO, Y. Rocky**; Occhiuti Rohlicek & Tsao LLP, 10 Fawcett Street, Cambridge, Massachusetts 02138 (US).

(22) International Filing Date:
3 September 2010 (03.09.2010)

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/239,534 3 September 2009 (03.09.2009) US

(71) Applicants (for all designated States except US):
ACADEMIA SINICA; 128, Sec 2, Academia Road, Nan-Kang, Taipei (TW). **LIANG, Chi-Ming** [—/US]; 6627 River Road, Bethesda, Maryland 20817 (US).

(72) Inventors; and

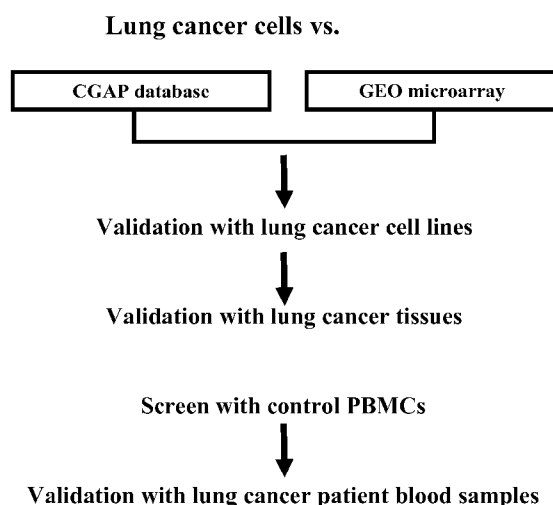
(75) Inventors/Applicants (for US only): **PECK, Konan**; 6F., No. 8 Alley 30, Lane 78, Sec. 2, Fu-Shin S. Road, Taipei City, 106 (TW). **HSEUH, Ming-Feng**; No. 123, Jhengyan Rd., Lingya District, Kaohsiung, 802 (TW). **SHIH, Jin-Yuan**; 3F., No. 7, Lane 231, Sec. 2, Jin-Shan

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, SE, SI, SK,

[Continued on next page]

(54) Title: MARKERS FOR DETECTING CANCER CELLS

FIG. 1



(57) Abstract: Detecting presence of cancer cells in a subject using one or more of fibronectin 1, S100 calcium binding protein A7, and carboxy-terminal domain (RNA polymerase II, polypeptide A) small phosphatase-like protein as a marker.



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG). **Published:**

— *without international search report and to be republished
upon receipt of that report (Rule 48.2(g))*

MARKERS FOR DETECTING CANCER CELLS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application No.
5 61/239,534, filed on September 3, 2009, the content of which is hereby
incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Carcinoma, developed from epithelial cells, accounts for around 85%
10 of human cancers. It has been reported that carcinoma cells exist in the bone
marrow and peripheral blood of cancer patients.

Methods such as immunocytology and flow cytometry have been
developed and employed in commercial instruments (e.g., CellSearch,
Veridex, NJ) to detect circulating tumor cells (CTC) in peripheral blood. See
15 US Patent No. 6,670,197; Budd, et al., *Clin. Cancer Res.*, 2006, 12:
6403-6409; and Riethdorf, et al., *Int. J. Cancer*, 2008, 123: 1991-2006).
These methods all use antibodies specifically targeting epithelial cell surface
markers (e.g., EpCAM) to detect carcinoma cells. Their sensitivity is limited
by the affinity of the antibody to the CTC cell surface marker it binds, and
20 they require a large amount of blood and yield a high rate of false negative
results. See Ring, et al., *Br. J. Cancer*, 2005, 92: 906-912.

Compared with immunocytology and flow cytometry, polymerase
chain reaction (PCR) is more facile and less expensive to perform. Peck, et
al. developed a scoring method, using quantitative PCR (qPCR), by
25 calculating cancer cell load (Lc) in peripheral blood sample (Peck, et al., U.S.
Patent No. 7,507,534) and the Lc value correlated well with the cell count of
CTC.

To detect CTCs at high sensitivity and high accuracy, new marker
genes as well as qPCR control genes are needed.

30

SUMMARY OF THE INVENTION

The present application is based on, at least in part, on the discovery of novel marker genes and a control gene for detecting the presence of circulating tumor cells (CTCs) in the blood of cancer patients.

5 Accordingly, provided herein are methods for detecting cancer cells in a subject (e.g., a human) or a sample. This methods include (i) providing a sample (e.g., a peripheral blood sample) of the subject that is suspected of containing cancer cells, (ii) determining in the sample the expression level of one or more of marker genes, (iii) determining in the sample the expression
10 level of a control gene, (iv) normalizing the expression level(s) of the one or more marker genes against that of the control gene to obtain an expression profile of the one or more marker genes, and (v) determining whether the sample contains cancer cells based on the expression profile. An expression profile representing an expression level of the one or more marker genes
15 higher than that in a cancer-free subject indicates that the sample contains cancer cells.

 The marker genes include fibronectin 1 (FN1), S100 calcium binding protein A7 (S100A7), and carboxy-terminal domain (RNA polymerase II, polypeptide A) small phosphatase-like protein (CTDSPL). In some cases,
20 the expression levels of all three marker genes in the sample are determined. In one example, the cancer cells to be detected are circulating lung cancer cells (e.g., non-small cell lung cancer cells). The expression levels of the marker genes and the control gene (e.g., DEAD box polypeptide 5; DDX5) can be determined by real-time quantitative PCR (qPCR) analysis.

25 The details of one or more embodiments of the invention are set forth in the description below. Other features or advantages of the present invention will be apparent from the following drawings and detailed description of an example, and also from the appended claims.

30 BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram depicting an exemplary screening process for

identifying marker genes useful in detecting circulating lung cancer cells ("CLCC") in the peripheral blood of lung cancer patients.

FIGs. 2A-C are graphs showing distribution of cancer cell load (Lc) values of marker genes for lung cancer patients and normal PBMC controls.

5 (A) FN1; (B) S100A7; and (C) CTDSPL.

FIGs. 3A and B are two graphs demonstrating positive CLCC detection rates using FN1, S100A7, and/or CTDSPL as markers. (A) positive CLCC detection rates using FN1, S100A7, or CTDSPL as a marker. (B) positive CLCC detection rates using a combination of FN1, S100A7, and
10 CTDSPL as a marker.

DETAILED DESCRIPTION OF THE INVENTION

Described herein are, *inter alia*, methods for detecting cancer cells, particularly circulating tumor cells, based on the expression level of one or
15 more of marker genes FN1, S100A7, and CTDSPL, normalized against that of a control gene (e.g., DDX5).

Circulating Tumor Cells

Circulating tumor cells (CTCs) are believed to be cells that have
20 detached from a primary tumor and then released into circulation. Studies suggest that analysis of CTCs can help monitoring and predicting cancer progression and evaluating response to treatments. See, e.g., Peck et al., *Cancer Research*, 1998, 58: 2761-2765; and Budd et al., *Clin Cancer Res*, 2006,12(21): 6403-6409.

25

Marker and Control Genes

Novel marker and control genes useful for detecting CTCs using quantitative PCR are described herein. Shown below are the gene sequences of FN1, S100A7, CTDSPL, and DDX5:

30

FN1 (GenBank Accession No.: NM_212482; SEQ ID NO:1):

```

1   gccccgcgcg gctgtgctgc acagggggag gagagggaa cccaggcgcg agcgggaaga
61  ggggacctgc agccacaact tctctgggtcc tctgcatccc ttctgtccct ccaccggtcc
5   121 ccttccccac cctctggccc ccaccttctt ggaggcgaca acccccgga ggcattagaa
181 gggatttttc ccgcaggttg cgaagggaa ccaacttggg ggcaacttgc ctcccggtgc
241 gggcgtctct cccccaccgt ctcaacatgc ttaggggtcc ggggccccgg ctgctgctgc
301 tggccgtcca gtgcctgggg acagcgggtgc cctccacggg agcctcgaag agcaagaggc
361 aggcctcagca aatggttcag cccagtgccc cgggtggctgt cagtcaaagc aagcccggtt
10  421 gttatgacaa tggaaaacac tatcagataa atcaacagtg ggagcggacc tacctaggca
481 atgcgtttgt ttgtacttgt tatggaggaa gccgaggttt taactgcgag agtaaacctg
541 aagctgaaga gacttgcttt gacaagtaca ctgggaacac ttaccgagtg ggtgacactt
601 atgagcgtcc taaagactcc actgtacctg catcggggtg ggcgaggga gggtcaggga
661 gaataagctg taccatcgca aaccgctgcc atgaaggggg tcagtcctac agattgggtg
15  721 acacctggag gagaccacat gagactggtg gttacatggt agagtgtgtg tgtcttggtg
781 atggaaaagg agaattggacc tgcaagccca tagctgagaa gtgttttgat catgctgctg
841 ggacttccta tgtggtcgga gaaacgtggg agaagcccta ccaagcgttg ccaagcgttg
901 attgtacttg cctgggagaa ggcagcggag gcatcacttg cacttctaga aatagatgca
961 acgatcagga cacaaggaca tcctatagaa ttggagacac ctggagcaag aaggataatc
20  1021 gaggaaacct gctccagtg c atctgcacag gcaacggccg agggagagtgg aagtgtgaga
1081 ggcagacctc tgtgcagacc gatctggccc ctccaccgat ctccaccgat gttcgcagag
1141 ctgtttacca accgcagcct cccccccagc ctctcccta tggccactgt gtcacagaca
1201 gtggtgtggt ctactctgtg gggatgcagt ggctgaagac acaaggaaat aagcaaattg
1261 tttgcacgtg cctgggcaac ggagtcagct gccaaagagac agctgtaacc cagacttacg
25  1321 ctggcacttc aaatggagag ccatgtgctc taccattcac ctacaatggc aggcgttctc
1381 actcctgcac cacagaaggg cgacaggacg gacatctttg gtgcagcaca acttcgaatt
1441 atgagcagga ccagaaatac tctttctgca cagaccacac tgttttggtt cagactcgag
1501 gaggaaatcc caatggtgcc ttgtgccact tccccttctc atacaacaac cacaattaca
1561 ctgattgcac ttctgagggc acatgaagtg gtgtgggacc acacagaact
30  1621 atgatgccga ccagaagttt ggggtctgcc ccatggctgc ccacgaggaa atctgcacaa
1681 ccaatgaagg ggtcatgtac cgcattggag atcagtgagg taagcagcat gacatgggtc
1741 acatgatgag gtgcacgtgt gttgggaatg gtcgtgggga atggacatgc attgcctact
1801 cgagcttctc agatcagtg c attgttgatg acatcactta caatgtgaac agcacattcc
1861 acaagcgtca tgaagagggg cacatgctga actgtacatg cttcggtcag ggtcggggca
35  1921 ggtggaagtg tgatcccgtc gaccaatgcc aggattcaga gactgggacg ttttatcaaa
1981 ttggagattc atgggagaag tatgtgcatg gtgtcagata ccagtgtctc tgctatggcc
2041 tggcatttgg ggaatggcat tgccaaactt tacagacctc tccaagctca agtggctctg
2101 tcgaagtatt tatcactgag actccgagtc agcccaactc ccaccctatc cagtggaaatg
2161 caccacagcc atctcacatt tccaagtaca ttctcaggtg gagacctaaa aattctgtag
40  2221 gccgttgga ggaagctacc ataccaggcc acttaaaact ctacaccatc aaaggcctga
2281 agctcgggtg ggtatacgag tgcagatcca tcagcatcca gcagtaagg caccagaag
2341 tgactcgctt tgacttcacc accaccagca ccagcacacc tgtgaccagc aacaccgtga
2401 caggagagac gactcccttt tctcctcttg tggccacttc tgaatctgtg accgaaatca
2461 cagccagtag ctttgtggtc tcctgggtct cagcttccga caccgtgtcg ggattccggg
45  2521 tggaaataga gctgagtgag gctgagtgag agccacagta cctggatctt ccaagcacag
2581 ccacttctgt gaacatccct gacctgcttc ctggccgaaa atacattgta aatgtctatc
2641 agatatctga ggaatggggg cagagtttga tcctgtctac ttcaaaaca acagcgctg
2701 atgccccctc tgacccgact gtggaccaag ttgatgacac ctcaattggt gttcgctgga
2761 gcagacccca ggctcccatc gaatagtcta ttcgccatc ttcgccatc atgaaggta
50  2821 gcagcacaga actcaacctt cctgaaactg caaactcgtg caccctcagt gacttgcaac
2881 ctggtgttca gtataacatc actatctatg ctgtggaaga aaatcaagaa agtacacctg
2941 ttgtcattca acaagaaacc actggcacc cagctcaga tacagtgcct tctcccaggg
3001 acctgcagtt tgtggaagtg aggtcaccat catgtggaca cctgtggaga ccgctgaga
3061 gtgcagtgac cggtaccgtg gtggatgtga tccccgtcaa cctgcctggc gagcacgggc
55  3121 agaggctgcc catcagcagg aacacotttg cagaagtcac cgggctgtcc cctgggggtc
3181 cctattactt caaagtcttt gcagtgagcc atgggagggg gagcaagcct ctgactgctc
3241 aacagacaac caaactggat gctcccacta acctccagtt tgtcaatgaa actgattcta
3301 ctgtcctggg gagatggact ccacctcggt cccagataac aggataccga ctgaccgtgg
3361 gccttaccgg aagaggacag ccaggcagat acaatgtggg tcctctgtgc tccaagtacc
60  3421 cactgaggaa tctgcagcct gcatctgagt acaccgtatc cctcgtggcc ataaagggca
3481 ccatccaagag ccccaaagcc actggagtct ttaccacact gcagcctggg agctctattc
3541 caccttacia caccgaggtg actgagacca ccatttgtat cacatggacg cctgctccaa
3601 gaattgggtt taagctgggt gtacgaccaa gccagggagg agaggcacca cgagaagtga
3661 cttcagactc aggaagcatc gttgtgtccc gcttgactcc aggagtagaa tacgtctaca
65  3721 ccatccaagt cctgagagat ggcaggaaa aattgttaac aattgttaac aaagtgggtg
3781 caccattgtc tccaccaaca aacttgcata tggaggcaaa ccctgacact ggagtgttca
3841 cagtctcctg ggagaggagc accaccccag acattactgg ttatagaatt accacaaccc
3901 ctacaaaacg ccagcaggga aattctttgg aagaagtggg ccatgctgat cagagctcct
3961 gcacttttga taacctgagt ccggcctgg agtacaatgt cagtgtttac cagtcaagg
70  4021 atgacaagga aagtgtccct atctctgata ccatcatccc agaggtgccc caactcactg

```

	4081	acctaagctt	tgttgatata	accgattcaa	gcatcggcct	gaggtggacc	ccgctaact
	4141	cttccaccat	tattgggtac	cgcacacag	tagttgcggc	aggagaaggt	atccctattt
	4201	ttgaagattt	tgtggactcc	tcagtaggat	actacacagt	cacagggctg	gagccgggca
5	4261	ttgactatga	tatcagcggt	atcactctca	ttaatggcgg	cgagagtgcc	cctactacac
	4321	tgacacaaca	aacgggtggt	cctcctccca	ctgacctgcg	attcaccaac	attgggtccag
	4381	acaccatgcg	tgtcacctgg	gctccacccc	catccattga	tttaaccaac	ttcctgggtgc
	4441	gttactcacc	tgtgaaaaat	gaggaagatg	ttgcagagtt	gtcaatttct	ccttcagaca
	4501	atgcagtggg	cttaacaaat	ctcctgcctg	gtacagaata	tgtagttagt	gtctccagtg
10	4561	tctacgaaca	acatgagagc	acacctctta	gaggaagaca	gaaaacaggt	cttgattccc
	4621	caactggcat	tgacttttct	gatattactg	ccaactcttt	tactgtgcac	tggattgtct
	4681	ctcgagccac	catcactggc	tacaggatcc	gccatcatcc	cgagcacttc	agtggggagac
	4741	ctcgagaaga	tcgggtgccc	cactctcgga	attccatcac	cctcaccaac	cctactccag
	4801	gtctacgata	tgtggtcagc	atcggtgcct	ttaatggcag	agaggaaagt	cccttattga
	4861	ttggccaaca	atcaacagtt	tctgatgttc	cgagggacct	ggaagtgtgt	gctgcgaccc
15	4921	ccaccagcct	actgatcagc	tgggatgctc	ctgctgtcac	agtgagatat	tacaggatca
	4981	cttacggaga	gacaggagga	aatagccctg	tccaggagtt	cactgtgcct	gggagcaagt
	5041	ctacaggata	catcagcgcc	cttaaacctg	gagttgatta	taccatcact	gtgtattccg
	5101	tacttgcccg	tggagacagc	cccgcaagca	gcaagccaat	ttccattaat	taccgaacag
	5161	aaattgacaa	accatcccag	atgcaagtga	ccgatgttca	ggacaacagc	attagtgtca
20	5221	agtggctgcc	ttcaagttcc	cttgttactg	gttacagagt	aaccaccact	cccaaaaatg
	5281	gtccaggacc	aacaaaaact	aaaactgcag	gtccagatca	aacagaaatg	actattgaag
	5341	gcttgccagcc	cacagtggag	tatgtgggtta	gtgtctatgc	tcagaatcca	agcggagaga
	5401	gtcagcctct	ggttcagact	gcagtaacca	acattgatcg	ccctaaagga	ctggcattca
	5461	ctgatgtgga	tgtcgattcc	atcaaaatgg	cttgggaaaag	cccacagggg	caagtttcca
25	5521	ggtagctggg	gacctactcg	agccctgagg	atggaatcca	tgagctattc	cctgcacctg
	5581	atggtgaaga	agacactgca	gagctgcaag	gcctcagacc	gggttctgag	tacacagtca
	5641	gtgtggttgc	cttgccagat	gatattggaga	gccagcccct	gattggaacc	cagtccacag
	5701	ctattcctgc	accaactgac	ctgaagtcca	ctcaggtcac	accacaagc	ctgagcgccc
	5761	atgggacacc	acccaatggt	cagctcactg	gatatcgagt	gcgggtgacc	cccaaggaga
30	5821	agaccggacc	aatgaaagaa	atcaaccttg	ctcctgacag	ctcatccgtg	gttgatcag
	5881	gacttatggg	ggccaccaaaa	tatgaagtga	gtgtctatgc	tcttaaggac	actttgacaa
	5941	gcagaccagc	tcagggagtt	gtcaccacta	tggagaatgt	cagcccacca	agaagggctc
	6001	gtgtgacaga	tgtctactgag	accaccatca	ccattagctg	gagaaccaag	actgagacga
	6061	tacttggtct	ccaagttgat	gccgttccag	ccaatggcca	gactccaatc	cagagaacca
35	6121	tcaagccaga	tgtcagaagc	tacaccatca	caggtttaca	accaggcact	gactacaaga
	6181	tctacctgta	caccttgaat	gacaatctca	ggagctcccc	tgtgggtcatc	gacgcctcca
	6241	ctgccattga	tgcaccatcc	aacctgcgtt	tccctggccac	cacaccaaat	tccctgctgg
	6301	tatcatggca	gccgccacgt	gccaggatta	ccggctacat	catcaagtat	gagaagcctg
	6361	ggtctcctcc	cagagaagtg	gtccctcgcc	cccgccctgg	tgtcacagag	gctactatta
40	6421	ctggccctgga	accgggaacc	gaataataca	tttatgtcat	tgccctgaag	aataatcaga
	6481	agagcgagcc	cctgattgga	aggaaaaaga	cagacgagct	tccccaaactg	gtaaaccttc
	6541	caacccccaa	tcttcatgga	ccagagatct	tggatgttcc	ttccacagtt	caaaagaccc
	6601	ctttcgtcac	ccaccctggg	tatgacactg	gaaatgggat	tcagcttcct	ggcacttctg
	6661	gtcaccctgc	cagtgttggg	caacaaatga	tctttgagga	acatggtttt	aggcggacca
45	6721	caccgcccac	aacggcccac	cccataaggc	ataggccaag	accatacccg	ccgaattgtag
	6781	gtgaggaaat	ccaaattggg	catatcccca	gggaagatgt	agactatcac	ctgtaccac
	6841	acggtccggg	actcaatcca	aatgcctcta	caggacaaga	agctctctct	cagacaacca
	6901	tctcatgggc	cccattccag	gacacttctg	agtacatcat	ttcatgtcat	cctgttggca
	6961	ctgatgaaga	acccttacag	ttcagggttc	ctggaacttc	taccagtggc	actctgacag
50	7021	gcctcaccag	aggtgccacc	tacaacatca	tagtggaggc	actgaaagac	cagcagaggc
	7081	ataaggttcg	ggaagaggtt	gttacccgtg	gcaactctgt	caacgaaggc	ttgaaccaac
	7141	ctacggatga	ctcgtgcttt	gacccctaca	cagtttccca	ttatgccgtt	ggagatgagt
	7201	gggaacgaat	gtctgaatca	ggctttaaac	tgttgtgcca	gtgcttaggc	tttgggaagtg
	7261	gtcatttcag	atgtgattca	tctagatggg	gccatgacaa	tggtgtgaac	tacaagattg
55	7321	gagagaagtg	ggaccgtcag	ggagaaaatg	gccagatgat	gagctgcaca	tgtcttggga
	7381	acggaaaagg	agaattcaag	tgtgaccctc	atgaggcaac	gtgttatgat	gatcggaaga
	7441	cataccacgt	aggagaacag	tggcagaagg	aatatctcgg	tgccatttgc	tccctgcacat
	7501	gctttggagg	ccagcggggc	tggcgctgtg	acaactgccg	cagacctggg	ggtgaaccca
	7561	gtcccgaagg	cactactggc	cagtcctaca	accagtatcc	tcagagatac	catcagagaa
60	7621	caaacactaa	tgttaattgc	ccaattgagt	gcttcatgcc	tttagatgta	caggctgaca
	7681	gagaagattc	ccgagagtaa	atcatctttc	caatccagag	gaacaagcat	gtctctctgc
	7741	caagatccat	ctaaaactgga	gtgatgttag	cagacccagc	ttagagttct	tctttctttc
	7801	ttaagccctt	tgtctgtggg	gaagttctcc	agcttcagct	caactcacag	cttctccaag
	7861	catcacccctg	ggagtttcc	gggggttttc	tcataaatga	gggctgcaca	ttgctgttcc
65	7921	tgtctcgaag	tattcaatac	cgtcagtat	tttaaatgaa	gtgattctaa	gatttgggtt
	7981	gggatcaata	ggaaagcata	tcagccaac	caagatgcaa	atgttttgaa	atgatatgac
	8041	caaaattttta	agtaggaaaag	tcaccccaac	acttctgctt	tcacttaagt	gtctggcccg
	8101	caatactgta	ggaacaagca	tgaattgttt	actgtgatat	tttaaatatc	caagatactc
	8161	actttttcca	aatgatcccta	gtaattgcct	agaaatatct	ttctcttacc	tgttatttat
70	8221	caatttttcc	cagtattttt	atacggaaaa	aattgtattg	aaaacactta	gtatgcagtt
	8281	gataagagga	atttgggtata	attatgggtg	gtgattattt	tttatactgt	atgtgccccaa

8341 gctttactac tgtggaaaga caactgtttt aataaaagat ttacattcca caacttgaag
 8401 ttcacatatt tgatataaga caccttcggg ggaaataatt cctgtgaata ttctttttca
 8461 attcagcaaa catttgaaaa tctatgatgt gcaagtctaa ttggtgattt cagtacaaga
 8521 ttttctaaat cagttgctac aaaaactgat tggtttttgt cacttcatct cttcactaat
 5 8581 ggagatagct ttacactttc tgccttaata gatttaagtg gaccccaata tttattaaaa
 8641 ttgctagttt accgttcaga agtataatag aaataatctt tagttgctct tttctaacca
 8701 ttgtaattct tcccttcttc cctccacctt tccttcattg aataaacctc tgttcaaaga
 8761 gattgctgctc aagggaata aaaatgacta agatattaaa aaaaaaaaaa aaaaa

10 S100A7 (GenBank Accession No.: NM_002963; SEQ ID NO:2)

1 gtccaaacac acacatctca ctcaccttc tactcgtgac gcttcccage tctggctttt
 61 tgaagcaaaa gatgagcaac actcaagctg agaggccat aataggcatg atcgacatgt
 121 ttcacaaata caccagacgt gatgacaaga ttgagaagcc aagcctgctg acgatgatga
 181 aggagaactt ccccaacttc cttagtgcct gtgacaaaaa gggcacaaat tacctcgccg
 15 241 atgtctttga gaaaaaggac aagaatgagg ataagaagat tgatttttct gagtttctgt
 301 ccttgctggg agacatagcc acagactacc acaagcagag ccatggagca gcgcctgtt
 361 ccgggggcag ccagtgacc agccccacca atgggcctcc agagacccca ggaacaataa
 421 aatgtcttct cccaccagaa aaaaaaaaaa

20 CTDSPL (GenBank Accession No.: NM_001008392; SEQ ID NO:3)

1 ggcggccgcg cgccgcgcac ccatggacgg cccggccatc atcaccaggg tgaccaaccc
 61 caaggaggac gagggccggt tgccggcgcc gggcgagaaa gcctcccagt gcaacgtcag
 121 cttaaagaag cagaggagcc gcagcatcct tagctccttc ttctgctgct tccgtgatta
 181 caatgtggag gcccctccac ccagcagccc cagtgtgctt ccgccactgg tggaggagaa
 25 241 tgggtgggctt cagaagggtg accagaggca ggtcattccc ataccaagtc caccagctaa
 301 gtaccttctt ccagagggtga cgtgcttga ctatggaaag aaatgtgtgg tctattgatt
 361 agatgaaaca ttggtgcaca gttcgtttaa gcctattagt aatgctgatt ttattgttcc
 421 ggttgaaatc gatggaacta tacatcaggt gtatgtgctg aagcggccac atgtggacga
 481 gttcctccag aggatggggc agctttttga atgtgtgctc ttactgccca gcttggccaa
 30 541 gtatgcagac cctgtggctg acctcctaga ccgctggggg gtgttccggg cccggctctt
 601 cagagaatca tgtgtttttc atcgtgggaa ctacgtgaag gacctgagtc gccttgggag
 661 ggagctgagc aaagtgatca ttgttgacaa tccccctgcc tcatacatct tccatcctga
 721 gaatgcagtg cctgtgcagt cctgggtcga tgacatgacg gacacggagc tgtggacct
 781 catccccctt tttgagggcc tgagccggga ggacgacgtg tacagcatgc tgcacagact
 35 841 ctgcaatagg tagccctggc ctctgcctgc ctccgcctg tgcactctgg aacctctggc
 901 ctcaggggag ctgcctgtcc tcagctccct gggagctgaa agtgaggata ctccgtgctc
 961 caggccacag ggtgaatgtg gccatgccta cctgttttgt ttttttaaga acagaaacaa
 1021 ctatttttaa agaactcttt taagaaattt cataaaggga catgcatttt actgggtttg
 1081 cttttcttaa aacataccaa aaaagaaaaa aatagaaaaa aaaaaaaaaa aagcttgatc
 40 1141 tctatcagac ttcttttcaa ctgtcctccc ctccaagcag accacctgtc cctttctatc
 1201 caagctcaga gcagctgacc caactcagaa tctctttcct acaggatgaa gtgccttttg
 1261 aatgtttatt taagccgaga gttaattttt ctacacaaca tatttccaga catcttttag
 1321 tcttttattg tcttagatac tataagaaga tgaacatgac aattttctag aacctggtag
 1381 cgtgtgtgtg tgtggcgggg ggtgctgagg gaggggagtg agtcacagga gcctgtcccc
 45 1441 gaacaggtgt gactgctctg acaactctgt gcatgctgca gggcaggct cagtgatgga
 1501 gcatctctat actatgtcat tctctccact catttttgac ccagtttggg atgtatctgc
 1561 aattgtgtgg ctcaacactt taggaaacaa tagattattt tatattatta tttctgatgg
 1621 tgacaagttt gtcttgaggt cacattttct ccttgaaaaa tgacatcctg tcaactctgc
 1681 tctcacacta ctgccataca ttgtgttttt ttgttgggta ttttgggta gagcagttac
 50 1741 aagaaaacct aaaacctttg gatataaaag aaatctgttt attgattttt aaatctttcc
 1801 tttccaaaag ctggatacac atggagctgt ttgggaattt tccttgctgc taccgcctg
 1861 ccaccaaatt gaattgacca gcggctgtta cactgttctt tgccactgtg cctatgctca
 1921 gaatagctc actgctaagc tacaaactcg gacaggggtc gaaacagag tgtccctcc
 1981 cattgcagcc tccaccacct gtaacctctt cctggcattg gccactgaag ggtacaaagg
 55 2041 caaaaggacc acagcaccac ttagggtgtg catggatttt aaactgcagt cagtatcaga
 2101 tcctgtttga taaaataagct gactgttctc tcttgagaac ctgtggcctc aaccagccac
 2161 caagctgatg tggcccaagt ccatctcttg gtcttctcct ttgaagcaca gcctattttc
 2221 gagccaaggg ttgggggaagc ctgtctagat gtgggactca ttgccccaaa ccaggagag
 2281 gaagagctcc cacagggaga gccaggctc tctttgcagc ctttccagb ttggtgttta
 60 2341 agcagtgcca tgttctttgt ttgacaacaa gacagctctg aaagtattgc tcttaaaaaa
 2401 aattaaaaag aaccctttca attggccacc attgccttag tctctgtgg gttggcttc
 2461 agccagcatt ctggtgggag tgactggcat taacaagact ggaaatcggg ggtcaaaagta
 2521 aaatatcttt gttttgcttt cattcacaaa gtaatgaagc cagctgccaa ttacatctc
 2581 ccaacagcac tttggctctgt ggactgctgt gtgaatattc agaaggggaag taagtattca
 65 2641 gggggtaaac aggtctccca gactctgag tgttccaaac cagtaattcc catgccaat
 2701 caaatagaac agccccctgc tagatattac cacagataat gacagttaca tggtagaact
 2761 gccatgcca caaatattta tttggaaaag tagtcattaa atgaaccac tgcttaaat
 2821 gtcttgaatg ttgcagtcaa gtgtctgtca tgtgttgata tccacacaga attaggccct

2881 aatgagagcc ttagaccctc aaccatgccc ccttcgttgg catcacaggg ccttatttgg
 2941 aagagcgggc aaagaggatg gaaatcataa aatatttcat gggaatcgaa cctagggata
 3001 gtgctccact tctgacgatg gagtgaagac acttggcaga cttgagccag acacttcacc
 3061 tagtagttcc tgaactgtg agcaccactg cactaagcca gtgcggagct gttaggggacg
 5 3121 ggcccagctc ctgcaccacg gacacagaat gtctggagag ggccagcagg ccctctgagg
 3181 gttctggaat ctgtgcacct ttttgacca cactccaaaa ttctgttttt attttaaccc
 3241 ttgaatctgc tttatgtaca taatcaaaat atctatatct atatctatat ctatatctat
 3301 atatttttaa tcatctacat gtaaataaag caatagaatt ctaacataag gccagaaat
 3361 aggacgaatg tttgggggtt atgtttttta aggtaaatac ggggtattgt ttttaattat
 10 3421 accatgtatt aaattgtggg ctttgaaacc taatgaaacc tgtagccac ttctctgtgc
 3481 catatacttc ccatgttacc aaaatacccc caactcttta gccaaaagag aaccctgacc
 3541 tccctgagtt ccctgtccct ttctgtacca ggtttaaatg tagtcttctg gagaagtatt
 3601 tttgacattg agctctggga caggacacct tgggtttgtg gactgcagc gactctgag
 3661 ttattacttc tctggccagg cctccagtgg aagtgcacag gcactcccaa tgtgttaat
 15 3721 gctctgtctt ccatttgttc tgggaatccta cgtgttggtc tgtggttcca tgcattagct
 3781 gtttgtaaat aatgcatttg catactgaaa aaggaatgcc acctgccaca gttgatggg
 3841 aggaagctcc tttgacgtgg tgcaattttg atgagatgtc tctggggaca cgaggatgcc
 3901 ctaatgatgc tgacttgtca tgggtgcagc atttgaactt ttggtgttaa aaaaaaaaaac
 3961 ctgtaagtct gtaacctggc aacattttac aaccctgtat ttttaaagat ggctttctaa
 20 4021 taaaaaatcc agaaccacac agccctatgg tcaacaatc ctacgtttgt gcctctgctt
 4081 ttaaagggtg tgtgctggac agttggcatg ccagggttcg agaagagtga atggcttgac
 4141 gtacttgca ttaactgtgc aaaattggct ggctgcctct gtccctactg tactgtaact
 4201 ttgatcatgt ctgttcctgt tccattctcc caggagcttc tctgcagact gacacaccct
 4261 cccccaccc gggtagtgga gatctgggt tctgggtagt catggatttc tgtggacat
 25 4321 ttgaatgtga taaacaatcc agcattactt aggaaatgct acatgaggaa tgtgcacgt
 4381 tccaggggag agtattgtca atcaaaaggt ttgcaatgat ttcttctctg ccaaaaaata
 4441 acatgtgaaa ctgca

DDX5 (GenBank Accession No.: NM_004396; SEQ ID NO:4)

30 1 atagagcggc tcccagcggt cctgcggcg taggagcgcg tccagactat aaaagcggt
 61 gccggaaaagc ggccggcacc tcattcattt ctaccggtct ctagttagtc agcttcggt
 121 ggtgtcatcg gtgtccttcc tccgctgcg ccccgcaag gcttcgcgt catcgaggcc
 181 atttccagcg acttgtcgca cgcttttcta tatacttcgt tcccggcaa ccgcaaccat
 241 tgacgccatg tgggttatt cgagtgcagc agaccgcggc cgggaccgag ggtttgggtg
 35 301 acctcgattt ggaggaagta gggcagggcc cttatctgga aagaagtttg gaaaccctgg
 361 ggagaaatta gttaaaaaga agtggaaatc tgatgagctg cctaaatttg agaagaatt
 421 ttatcaagag caccctgatt tggctaggct cacagcaca gaggtgaaa cacaagaag
 481 aagcaaggaa attacagtta gaggtcaca ctgcccgaag ccagttctaa atttttatga
 541 agccaatttc cctgcaaatg tcatggatgt tattgcaaga cagaatttca ctgaaccac
 40 601 tgctattcaa gctcagggat ggccagtgtc tctaagtgga ttggatatgg ttggagtggc
 661 acagactgac tctgggaaaa cattgttctc ttgtcttct gccattgtcc acatcaatca
 721 tcagccattc ctagagagag gcgatgggccc tatttgtttg gtgctggcac caactcggga
 781 actggcccaa caggtgcagc aagtagctgc tgaatattgt agagcatgtc gcttgaagtc
 841 tacttgtatc tacgggtgtg ctccaaagg accacaaata cgtgatttgg agagaggtgt
 45 901 ggaaatctgt attgcaacac ttgtaagctt gattgacttt ttagagtgtg gaaaaaccaa
 961 tctgagaaga acaacctacc ttgtccttga tgaagcagat agaatgcttg atatgggctt
 1021 tgaaacccaa ataaggaaga ttgtggatca aataagacct gataggcaaa ctctaattgt
 1081 gagtgcgact tggccaaaag aagtaagaca gcttgcgtga gatttctcta aagactatat
 1141 tcatataaac attggtgcac ttgaaactgag tgcaaacac aacattcttc agattgtgga
 50 1201 tgtgtgtcat gacgtagaaa aggatgaaaa acttattcgt ctaatggaag agatcatgag
 1261 tgagaaggag aataaaacca ttgtttttgt ggaaaccaa agaagatgtg atgagcttac
 1321 cagaaaaatg aggagagatg ggtggcctgc catgggtatc catgggtgac agagtcaaca
 1381 agagcgtgac tgggttctaa atgaattcaa acatggaaaa gctcctattc tgattgctac
 1441 agatgtggcc tccagagggc tagatgtgga agatgtgaaa tttgtcatca attatgacta
 55 1501 ccctaactcc tcagaggatt atattcoatg aattggaaga actgctcgca gtacaaaaac
 1561 aggcacagca tacactttct ttacacctaa taacataaag caagtgcagc accttatctc
 1621 tgtgcttcgt gaagctaatac aagcaattaa tcccaagttg cttcagttgg tcgaagacag
 1681 aggttcaggt cgttccaggg gttagaggag catgaaggat gaccgtcggg acagatactc
 1741 tgcgggcaaa aggggtggat ttaatacctt tagagacagg gaaaattatg acagaggtta
 60 1801 ctctagcctg cttaaaagag attttggggc aaaaactcag aatgggtgtt acagtgtgc
 1861 aaattacacc aatgggagct ttggaagtaa ttttgtgtct gctggatatac agaccagttt
 1921 taggactggg aatccaacag ggacttacca gaatggttat gatagcactc agcaatacgg
 1981 aagtaatgtt ccaaatatgc acaatgggat gaaccaacag gcatatgcat atcctgctac
 2041 tgcagctgca cctatgattg gttatccaat gccaacagga tattcccaat aagactttag
 65 2101 aagtatatgt aaatgtctgt ttttccaat tgctctttat attgtgtgtt atctgacaag
 2161 atagttatth aagaaacatg ggaattgcag aaatgactgc agtgcagcag taattatggg
 2221 gacttttttc gctattttaag ttggatattt ctctacattc ctgaaacaat ttttaggttt
 2281 tttttgtact agaaaatgca ggcagtgttt tcacaaaagt aaatgtacag tgatttgaaa
 2341 tacaataaat gaaggcaatg catggccttc caataaaaaa tatttgaaag ctgaattaa
 70 2401 tggaaattgt actttatttt atataatgtc atgtaaaact ttgcttaaga tgggtctgggt

2461 ttttttttggg ttttggttggg tttttttttt ccatgaaaac aaatgactgt tcctttttat
 2521 ttaatttggg aggcaggggg aatcagaagg cccttcttta taatgagcta ttcattattgc
 2581 aggagtcaga atgaattgat acaggtgaat ttttagttac aggcataaatt gcataaaaagc
 2641 tttgtcagct tccagcatca ggggagtcac ttaatagcct ttttccttat ttgctagtat
 5 2701 ggtttaaata gaaaatagta aaatagatac aaagtcacat atatagtgtg agaactggg
 2761 tgactttttc aaagtattata atttaaaaag ctccaaataa ctggcctttt caagagactt
 2821 atactcatgc tcttggctat actgtgaatt actgaaatgt tgaacaaacc tgtgaaagac
 2881 atacattagc cctttaagat ggccaggagc taagcttgag tctcctttac tgaatttcgt
 2941 tcttagtgca gggtacttgt agattctagt cttcacaggc tccctggggc tcttaactag
 10 3001 tcacactggg agtcatgaat gtctttccaa taattcaggg aattctagag atcctcaaac
 3061 tgtaagggtct attcatactc aacacaagga aaaaacctca ttaaaattaa tgactaatca
 3121 ggaggcaacg taaccaaaag cacagtgaat gaaagttttc atggtagggt caacatgggt
 3181 ttatttgctag aaagatccag gggatagcct taggtttaac ttcgggtcac caacgtaact
 3241 ttctaatacat ttatttcagt aatagctaga agtgggtctg aatgttttcc cagagtctga
 15 3301 tacgtgtttt tttttgccag aagagagggtc ttcaggagac ttcattttaa ttctgattat
 3361 taaactgagg ctttaattga tgttaatgcc ttatgtcaaa tgtaaagtta gaatttgcta
 3421 gggctgggat agggagtgat atttctagga cttagacatt gaaaactaat tcagcctgta
 3481 gtaacctgga tgggtttcaa tggcatgggt agtcaaattc atggttttaa acttagaagc
 3541 agctttcggg ggagagggtg ggttggagca tttattacat attttactgt ttaattgtctt
 20 3601 aaccgtgggc cttttaattt gtaaacactg aaatgattgt tgggctgtg aaaaacttta
 3661 cctattttacc ttggaagttt taaaagacag tccacttttt agcatgtgtg ttgtgtccag
 3721 cctgtgggtcg tcttaactaa taaatgtgat ttttctccca aaaaaaaaa

25 Diagnostic Methods

The present methods can be used to detect the presence of CTCs in a subject (e.g., a human subject) suspected or diagnosed as having cancer.

A test sample suspected of or known as containing tumor cells (e.g., a blood sample, a bone marrow sample, or a lymph node sample) is obtained
 30 from a human subject and the expression level of one or more marker genes FN1, S100A7, and CTDSPL, as well as that of a control gene can be determined by conventional methods. See, e.g., McDoniels-Silvers et al., Clin. Cancer Res, 2002, 8:1127-1138; Fleischhacker et al., Ann. N. Y. Acad. Sci., 2001, 945:179-188. It may be advantageous to determine the
 35 expression levels of all three marker genes in the same sample, as the data described herein suggest that accurate detection of CTCs is increased when the expression of all three marker genes is taken into account.

In one example, the expression level of a marker gene is determined by real-time quantitative PCR using a kit containing a set of primers specific to
 40 the one or more of the marker genes and the control gene. Alternatively, those of ordinary skill in the art would be able to design primers specific to the marker and control genes useful for real-time quantitative PCR.

The data indicating the expression level of a marker gene is first normalized against that of a control gene and the normalized data is then

processed by, e.g., computational analysis, to generate an expression profile that characterizes the expression level of the marker gene. This profile is compared with an expression profile representing the expression level of the same marker gene, normalized against that of the same control gene, in a cancer-free human subject. If the expression profile obtained from the test sample indicates a higher expression level of the marker gene than that of the cancer-free patient, it indicates that the sample contains cancer cells. Conventionally statistical analysis (e.g., the *t* test) can be employed to determine whether a sample from a particular subject (e.g., one who is suspected of having cancer) has a meaningfully higher expression level of a marker gene as compared to that of a sample from a cancer-free subject.

The relative expression level of a marker gene (normalized against that of a control gene) in peripheral blood can be calculated following the formula below:

$$\Delta C_{T(\text{marker})} = C_T^{(\text{control gene})} - C_T^{(\text{marker gene})}$$

 $C_T^{(\text{control gene})}$ and $C_T^{(\text{marker gene})}$ refer to the qPCR threshold cycle numbers of the control gene and the marker gene, respectively. If the fluorescence signal of a marker gene were not detected after 55 cycles, 55 was assigned as the C_T value of that marker gene.

The differential expression level (a test sample versus a healthy control sample) of a marker gene (Q) can be calculated as follows:

$$Q = \Delta C_{T(\text{test})} - (\text{mean of } \Delta C_{T(\text{healthy control})}),$$

in which $\Delta C_{T(\text{test})}$ and (mean of $\Delta C_{T(\text{healthy control})}$) refer to qPCR differential threshold cycle numbers obtained for a candidate cancer patient (a test sample) and the mean differential threshold cycle numbers obtained from healthy controls, respectively.

The normalized differential expression of marker gene j in test sample i (E_{ij}) can be calculated by the equation $E_{ij} = \frac{Q_{ij} - \bar{Q}_j}{\sigma_j}$, in which Q_{ij} refers to Q of marker gene j in test sample i ; $\bar{Q}_j$ refers to the mean of Q of marker gene

j in healthy controls; and σ_j refers to the standard deviation of Q of marker gene j in the healthy controls.

The load of cancer cells in the circulation, L_c , which has been shown to closely correlate with CTC numbers (see Peck et al., U.S. Patent No.

5 7,507,534), can also be calculated. A higher L_c score obtained from the sample of a suspected cancer patient (as compared to the L_c score obtained from the sample of a cancer-free subject) is indicative of the presence of cancer cells in the patient. L_c is defined as the sum of the E values of all markers:

10
$$L_c = \sum_{j=1}^n E_j, \text{ where } n \text{ is the number of marker genes.}$$

The method described herein can be used to detect the presence of tumor cells (e.g., lung cancer cells, breast cancer cells, colorectal cancer cells, and prostate cancer cells), particularly CTCs, in a patient. The present methods are useful for diagnosing cancer in patients suspected of having
15 cancer. For those patients already diagnosed with cancer, the present methods can be used to stage the cancer, monitor cancer progression, detect relapse and evaluate response to treatments.

Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present invention to its fullest
20 extent. The following specific example is, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. All publications cited herein are incorporated by reference.

EXAMPLE

25 This example describes the identification and verification of novel marker and control genes for detecting CTCs.

Materials and Methods

Peripheral blood samples were obtained from 45 patients with
30 histologically documented non-small cell lung cancer (NSCLC). These

patients were treated with chemotherapy. 21 healthy volunteers free of cancer were chosen as normal controls. Two blood samples were collected from each subject. The first 1-2 ml of peripheral blood was collected in the Vacutainer[®] (Becton Dickinson, Rutherford, NJ) tube containing EDTA, which was later discarded, and a subsequent 5 ml of blood in another Vacutainer[®] blood collection tube. Total RNAs were extracted from the 5 mo blood using the QIAamp RNeasy Blood Mini kit (Qiagen, Hilden, Germany) within 2 hours after the blood samples were collected. The RNA amount was quantified by UV/VIS spectrophotometer and the RNAs were stored in a -80°C freezer. The first-strand cDNA was generated as follows. 1 to 2 µg of the total RNAs was incubated with 2.5 µM oligo (dT)₁₈ primer at 70°C for 5 min and then chilled immediately on ice for 5 min. The reverse transcription reaction was conducted by adding 4 µl of 5× First-Strand Buffer, 5 mM DTT, 0.5 mM dNTP, 200 units of SuperScript[®] III reverse transcriptase, and 40 units of RNaseOUT[™] Recombinant RNase Inhibitor (Invitrogen, Carlsbad, CA) to the aforementioned RNA and primer mix solution to a final volume of 20 µl and allowed the reverse transcription reaction to proceed for 1.5 h at 50°C. The enzyme activity was inactivated by heating at 70°C for 20 min.

Real-time quantitative PCR (qPCR) was used to determine the amount of a CTC marker gene transcript. The qPCR primers and probes were selected from the Universal ProbeLibrary (UPL; Roche Diagnostics, Penzberg, Germany). qPCR reactions were performed using the LightCycler 480[®] (Roche Diagnostics, Penzberg, Germany). Reaction mixtures contained 5 µl of 2× Probes Master, 100 nM of UPL and 200 nM of primer (each) in a total volume of 10 µl. The qPCR conditions were: 95°C for 10 min, followed by 60 cycles at 95°C for 10 s, 60°C for 20 s, and 72°C for 2 s. Data thus obtained were analyzed by LC480 software.

30 Results

Identification of DDX5 as a control for detecting CLCC by qPCR analysis

The GAPDH gene is often used in qPCR assay as a control to measure the differential threshold cycle number, ΔC_T , between the control and the target genes to account for the variations in the sample processing procedures. Other than GAPDH, housekeeping gene such as ACTB (β -actin), 18S
 5 ribosomal RNA, and several others have been used as control genes in qPCR assay. Others, such as DEAD (Asp-Glu-Ala-Asp) box polypeptide 5 (DDX5), has been reported to serve as a control in analyzing gene expression profiles of cancer cell lines, cancerous tissues, and adjacent normal tissues. See Su, et al., *BMC Genomics*, 2007, 8: 140. It is unknown what control
 10 gene can be used in qPCR analysis to identify CTCs from a large pool of mononuclear cells in peripheral blood.

To identify control genes for CTCs, three genes, GAPDH, β -actin (ACTB), and DDX5, were tested for their use as controls in detecting CLCC. Their expression levels in 28 lung cancer patients and 16 healthy controls were
 15 determined by qPCR. Table 1 below shows the coefficient of variation (CV) of the expression levels of these genes in the cancer patients relative to those in the healthy controls.

20 Table 1. Expression Level Variations of GAPDH, ACTB, and DDX5 in Lung Cancer Patient Versus Healthy Controls

qPCR control gene	GAPDH	ACTB	DDX5
Coefficient of variation	1.69%	1.68%	1.2%
<i>P</i> value of unpaired t test	0.92	0.13	0.03

Among the three control genes, the CV of DDX5 (1.21%) is lower than GAPDH (1.69%) and ACTB (1.68%), indicating that the DDX5 expression
 25 level is more consistent in both lung cancer patients receiving chemotherapy and healthy volunteers.

To further test which control gene among the three yields better results in identifying CTCs from the overwhelmingly large number of PBMCs in blood plasma, we examined the expression level of CTDSPL and normalized
 30 it with that of GAPDH, ACTB, or DDX5 in both healthy control and lung

cancer patient samples. The *P* values obtained from unpaired t test shown in Table 1 indicate that DDX5 is the best internal control gene to be used in a qPCR assay for detecting CLCCs. The DDX5-specific primers used in this study are listed in Table 2 below.

5

Identification and verification of marker genes useful in detecting CLCCs

In the circulation system, the number of peripheral blood mononuclear cells (PBMC) is 5 to 6 times greater than that of circulating tumor cells in the peripheral blood. It is challenging to identify marker genes that are
10 expressed in tumor cells but are not expressed or are expressed at a much lower level in blood cells. Keratin 19 has been employed as a marker for CTC detection. The marker facilitates a detection limit of 1 CTC in 1 ml of blood and a positive detection rate around 50% (Peck, et al., Cancer Res., 1998, 58: 2761-2765).

15 To identify novel marker genes, candidate marker genes were first selected based on the Cancer Genome Anatomy Project (CGAP) database (an expressed sequence tag database) provided by the National Cancer Institute and the Gene Expression Omnibus (GEO) microarray database (an gene expression database) provided by the National Center for Biotechnology
20 Information. FIG. 1 shows the validation and screening process used in this example for identifying candidate marker genes. Genes having differential expression ratios between cancerous lung tissues/cancer cell lines and normal leukocytes were selected from the databases as candidate marker genes. These selected genes were further compared with literature reports to exclude
25 those that have already been reported for use as markers in detecting CLCCs (see Sheu, et al., *Int. J. Cancer*, 2006, 119: 1419-1426; Mitas, *J. Mol. Diagn.*, 2003, 5: 237-242). The remaining candidate marker genes were analyzed by qPCR to determining their expression levels in lung cancer cell lines and in lung cancer tissues. Those having high expression levels in cancer
30 cells/tissues were further analyzed by qPCR to determining their expression levels in peripheral blood mononuclear cells (PBMCs). 66 candidate marker

genes were identified as exhibiting high expression levels in cancer cells and low expression levels in PBMCs. Among them, three marker genes, fibronectin 1 (FN1), S100 calcium binding protein A7 (S100A7), and CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like (CTDSPL) were found to be useful in detecting CLCCs from lung cancer patients. The primers and probes for determining the mRNA levels of these three genes and the control gene are listed in Table 2 below:

Table 2. The primer and probe sequences specific to FN1, S100A1, CTDSPL, and DDX5

Gene		Sequence
FN1	Forward	gcaggtggaagtgtgatcc (SEQ ID NO:5)
	Reverse	agcagtagcactggtatctgaca (SEQ ID NO:6)
	Probe*	gggagaag (SEQ ID NO:7)
S100A7	Forward	ctgctgacgatgatgaagga (SEQ ID NO:8)
	Reverse	cgaggtaattgtgcccttt (SEQ ID NO:9)
	Probe*	tggggaag (SEQ ID NO:10)
CTDSPL	Forward	aaggacctgagtcgccttg (SEQ ID NO:11)
	Reverse	tccaaccaggactgcaca (SEQ ID NO:12)
	Probe*	gggagctg (SEQ ID NO:13)
DDX5	Forward	cagactggatctgggaaaaca (SEQ ID NO:14)
	Reverse	gtgccagcaccaacaaat (SEQ ID NO:15)
	Probe*	gcaggaag (SEQ ID NO:16)

*Locked nucleic acid (LNA) sequence

Semi-quantification of CTCs and statistical analysis

The relative expression level of a marker gene (normalized against that of a control gene) in peripheral blood was calculated following the formula below:

$$\Delta C_{T(\text{marker})} = C_T^{(\text{DDX5})} - C_T^{(\text{marker gene})}$$

$C_T^{(\text{DDX5})}$ and $C_T^{(\text{marker gene})}$ refer to the qPCR threshold cycle numbers of DDX5 and the marker gene, respectively. If the fluorescence signal of a marker gene were not detected after 55 cycles, 55 was assigned as the C_T value of that marker gene.

The differential expression level (a test sample versus a healthy control

sample) of a marker gene (Q) is calculated as follows:

$$Q = \Delta C_{T(\text{test})} - (\text{mean of } \Delta C_{T(\text{healthy control})}),$$

in which $\Delta C_{T(\text{test})}$ and (mean of $\Delta C_{T(\text{healthy control})}$) refer to qPCR differential threshold cycle numbers obtained for a candidate lung cancer patient (a test sample) and the mean differential threshold cycle numbers obtained from healthy controls, respectively.

The normalized differential expression of marker gene j in test sample

i (E_{ij}) is calculated by the equation $E_{ij} = \frac{Q_{ij} - \bar{Q}_j}{\sigma_j}$, in which Q_{ij} refers to Q of

marker gene j in test sample i ; $\bar{Q}_j$ refers to the mean of Q of marker gene j in

healthy controls; and σ_j refers to the standard deviation of Q of marker gene j in the healthy controls.

The load of cancer cells in the circulation, L_c , which has been proven to closely correlated with CTC numbers (see Peck et al., US Patent 7,507,534), is defined as the sum of the E values of all markers:

$$L_c = \sum_{j=1}^n E_j, \text{ where } n \text{ is the number of marker genes.}$$

Verification of the markers and enhancement of positive detection rate with multiple marker genes

As shown in Figs 2A-C, FN1, CTDSPL, and S100A7 exhibited higher L_c scores in lung cancer patients than in healthy controls and the results were statistically significant ($P < 0.05$). Determined by the t test, the P values of FN1, CTDSPL, S100A7 are 2.6E-02, 2.8E-02, and 6.5E-03, respectively.

The positive CLCC detection rates using these three marker genes individually are 47% (FN1), 64% (S100A7), and 64% (CTDSPL) in NSCLC patients.

See Fig. 3A. As shown in Fig. 2B, the positive CLCC detection rate based on CTDSPL alone is 64%, based on the combination of CTDSPL and S100A7 is 87%, and based on all three marker genes is 96%. Figs. 3A and 3B also show that positive detection increases with the number of the marker genes, i.e., from 64% (CTDSPL alone) to 87% (CTDSPL and S100A7) and 96%.

OTHER EMBODIMENTS

All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be
5 replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

From the above description, one skilled in the art can easily ascertain the essential characteristics of the present invention, and without departing
10 from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

What is claimed is:

1. A method for detecting cancer cells in a subject, comprising
5 providing a sample of the subject that is suspected of containing cancer cells,
determining in the sample an expression level of one or more marker genes selected from the group consisting of fibronectin 1 (FN1), S100 calcium binding protein A7 (S100A7), and carboxy-terminal domain (RNA
10 polymerase II, polypeptide A) small phosphatase-like protein (CTDSPL),
determining in the sample an expression level of a control gene,
normalizing the expression level of the one or more marker genes against that of the control gene to obtain an expression profile of the one or more marker genes, and
15 determining whether the sample contains cancer cells based on the expression profile;
wherein an expression profile representing an expression level of the one or more marker genes higher than that in a cancer-free subject indicates that the subject has cancer cells.
20
2. The method of claim 1, wherein the sample is a peripheral blood sample.
3. The method of claim 1, wherein the sample is suspected of
25 containing lung cancer cells.
4. The method of claim 3, wherein the lung cancer cells are non-small cell lung cancer cells.
5. The method of claim 1, wherein the expression levels of FN1,
30 S100A7, and CTDSPL are determined.

6. The method of claim 5, wherein the expression levels of FN1, S100A7, and CTDSPL are determined by real-time quantitative PCR (qPCR).

5 7. The method of claim 1, wherein the control gene is DEAD box polypeptide 5 (DDX5).

8. The method of claim 7, wherein the expression level of DDX5 is determined by real-time qPCR.

10

9. The method of claim 5, wherein the control gene is DDX5.

10. The method of claim 9, wherein the expression level of DDX5 is determined by real-time qPCR.

15

11. The method of claim 2, wherein the sample is suspected of containing lung cancer cells.

12. The method of claim 11, wherein the expression levels of FN1, S100A7, and CTDSPL are determined.

20

13. The method of claim 12, wherein the expression levels of FN1, S100A7, and CTDSPL are determined by real-time quantitative PCR (qPCR).

14. The method of claim 11, wherein the control gene is DEAD box polypeptide 5 (DDX5).

25

15. The method of claim 14, wherein the expression level of DDX5 is determined by real-time qPCR.

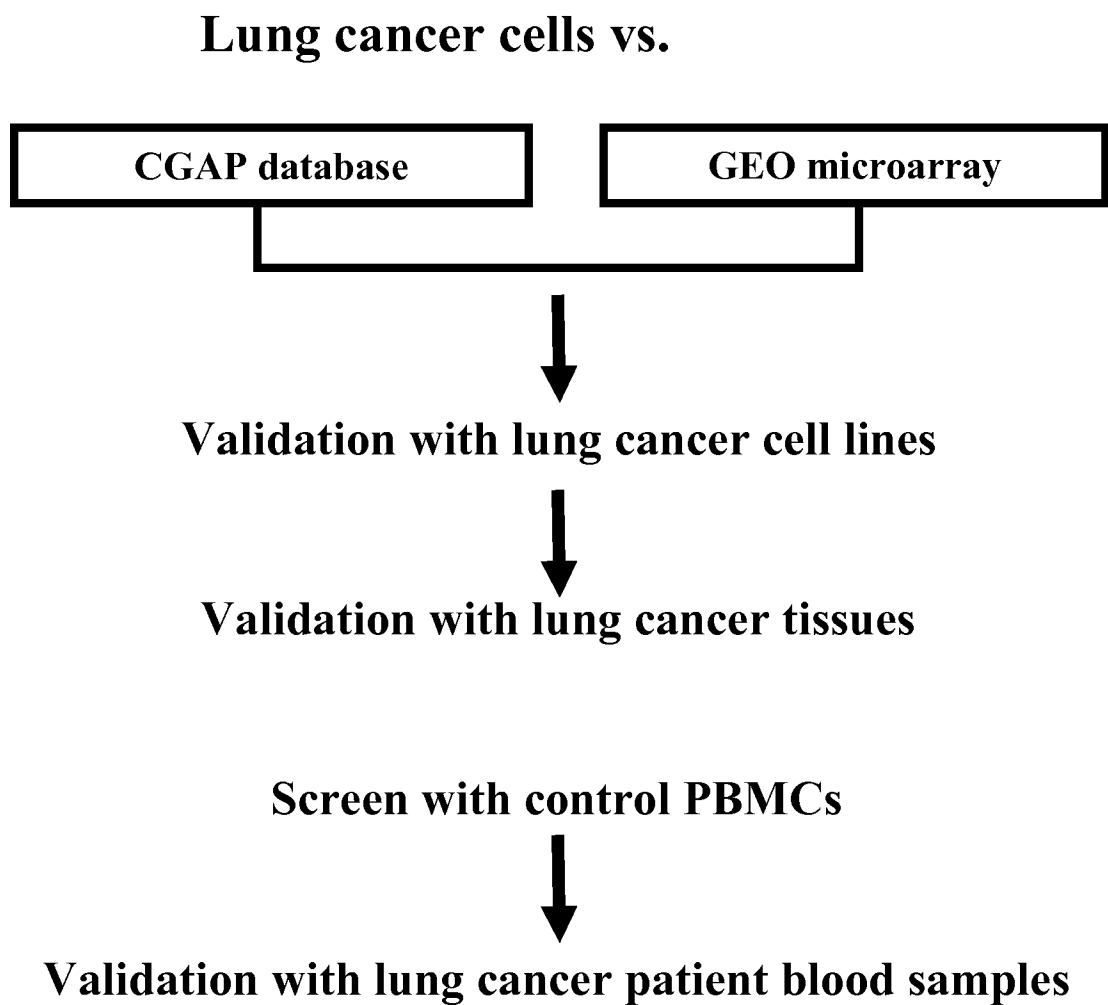
30

16. The method of claim 12, wherein the control gene is DDX5.

17. The method of claim 13, wherein the control gene is DDX5 and its expression level is determined by real-time qPCR.

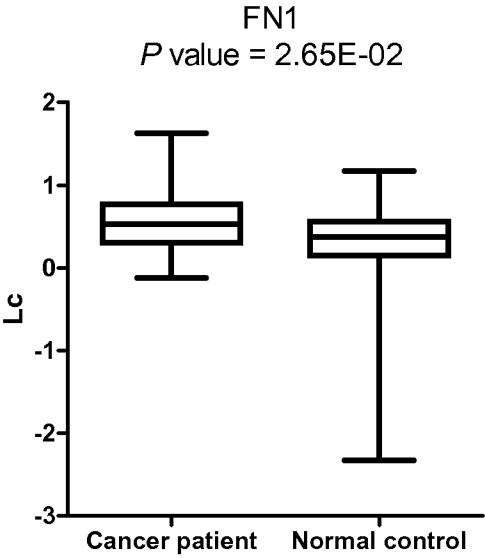
18. The method of claim 6, further comprising calculating a cancer
5 cell load (Lc) in the sample.

19. The method of claim 13, further comprising calculating a cancer cell load (Lc) in the sample.

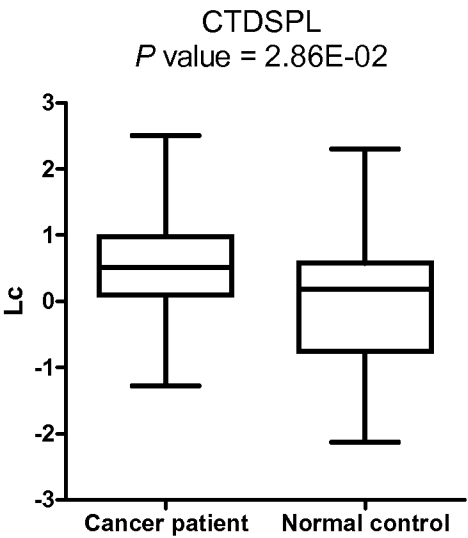
FIG. 1

FIGs. 2A-C

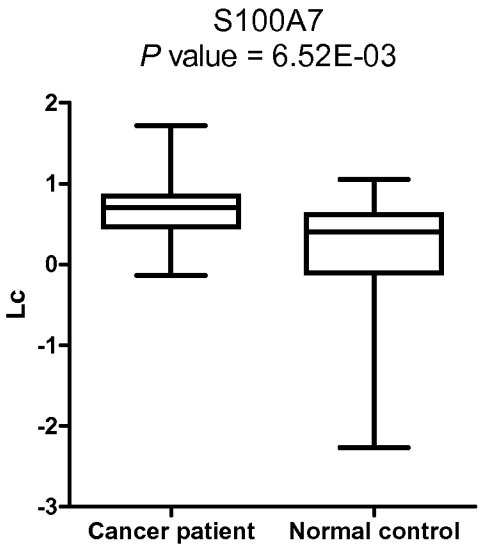
A.

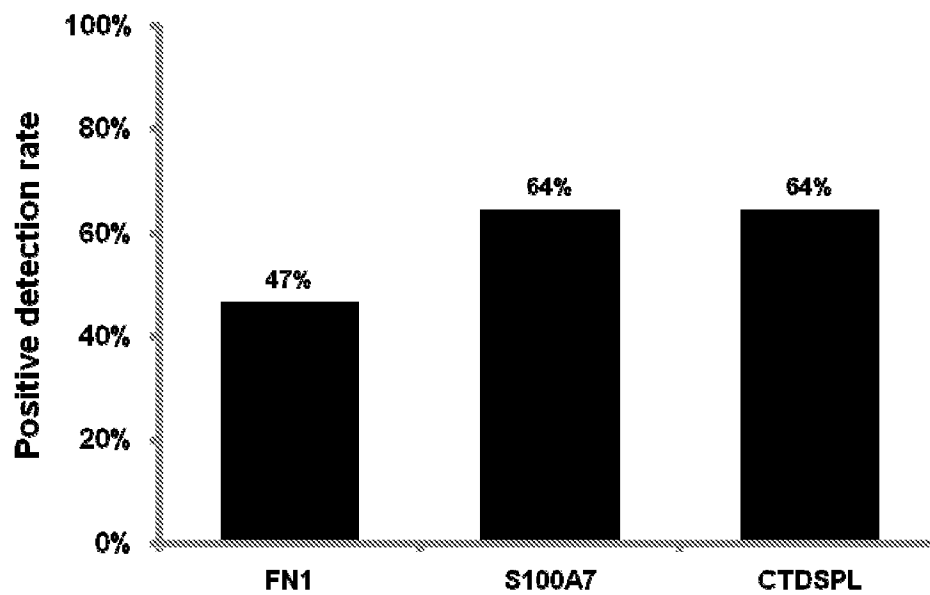


B.



C.



FIGs. 3A-B**A.****B.**