

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 July 2011 (07.07.2011)

(10) International Publication Number
WO 2011/079846 A2

PCT

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/DK2010/050358

(22) International Filing Date:
23 December 2010 (23.12.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 2009 70305 30 December 2009 (30.12.2009) DK

(71) Applicants (for all designated States except US):
RIGSHOSPITALET [DK/DK]; Blegdamsvej 9,
DK-2100 Copenhagen Ø (DK). **HERLEV HOSPITAL**
[DK/DK]; Herlev Ringvej 75, DK-2730 Herlev (DK).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ROSSING, Maria**
[DK/DK]; Smutvej 1, DK-2920 Charlottenlund (DK).
NIELSEN, Finn Cilius [DK/DK]; Vordingborggade 6E,
lejl. 33, DK-2100 Copenhagen Ø (DK). **BENNEDBÆK,**
Finn Noe [DK/DK]; Hjortevænget 66, DK-2880
Bagsværd (DK). **BORUP-HELWEG-LARSEN, Rehan-**
nah Holga Andrea [DK/DK]; Apollovej 15, DK-2720
Vanløse (DK).

(74) Agent: **HØIBERG A/S**; St. Kongensgade 59 A,
DK-1264 Copenhagen K (DK).

(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.

(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, RS, SE, SI, SK,
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))

(54) Title: MRNA CLASSIFICATION OF THYROID FOLLICULAR NEOPLASIA

(57) Abstract: The present invention relates to molecular classifiers based on specific m RNA expression patterns which distinguish the malignant and benign subtypes of thyroid follicular neoplasia; follicular thyroid carcinoma and follicular thyroid adenoma. The invention further relates to methods for diagnosing thyroid follicular neoplasia in samples from thyroid nodules to reduce the number of diagnostic operations and expediting surgery for individuals with a malignant nodule.



WO 2011/079846 A2

mRNA classification of thyroid follicular neoplasia

This application is a non-provisional application of DK patent application PA 2009 70305 filed on December 30, 2009, which is hereby incorporated by reference in its entirety. All patent and non-patent references cited in the application are hereby incorporated by reference in their entirety.

Field of invention

The present invention relates to molecular classifiers based on specific mRNA expression patterns which distinguish the malignant and benign subtypes of thyroid follicular neoplasia. The invention further relates to methods for differentiating between malignant and benign subtypes of thyroid follicular neoplasia. This can prove as a valuable pre-operative diagnostic tool; thus reducing the number of diagnostic operations and expediting surgery for individuals with a malignant nodule.

Background of invention

Thyroid nodules are a common clinical finding (Hegedus *et al.* 2003; Hegedus 2004). In Western Europe approximately 6% of all women have palpable nodules and the number of silent thyroid nodules is several fold higher. In addition to alleviate local compressive symptoms or thyroid hyperfunction, the major clinical challenge is to exclude the possibility of malignancy. Only about 5% of cold thyroid nodules become malignant and it is therefore important that the diagnostic procedures exhibit a high sensitivity and specificity (Ruggeri *et al.* 2008; Utiger 2005). Follicular thyroid carcinomas (FC) comprise about 15 % of all malignant nodules and they may be overlooked, since the diagnosis mainly relies on the exclusion of capsular and /or vascular invasion. It is moreover difficult to distinguish benign follicular adenoma (FA) from carcinoma.

The road to follicular neoplasia is not completely understood. In contrast to the well defined *RET* and *BRAF* mutations found in medullary and papillary thyroid cancer, follicular tumors do not exhibit consistent mutations (Fagin & Mitsiades 2008) although individuals exhibiting variations in *FOXE1* (*TTF2*) and *NKX2-1* (*TTF1*) have recently been reported to have increased risk of developing follicular carcinoma (Gudmundsson *et al.* 2009; Landa *et al.* 2009). *P53* and *RAS* are mutated in about half of the tumors and a recurrent *PAX8-PPAR γ* translocation has been identified in about 60% of the

- cancers, but also in a number of adenoma (Nikiforova *et al.* 2003). The microfollicular adenomas or fetal adenoma (FEA) represent a distinct but highly related type of follicular nodule exhibiting a high degree of aneuploidy, that renders these tumors more likely to become malignant (Castro *et al.* 2001). A number of studies have successfully exploited global expression profiling to identify molecular markers or signatures of thyroid neoplasia (Barden *et al.* 2003; Finley *et al.* 2004; Mazzanti *et al.* 2004; Lubitz *et al.* 2005; Weber *et al.* 2005; Fryknas *et al.* 2006; Griffith *et al.* 2006; Fijarewicz *et al.* 2007; Prasad *et al.* 2008; Hinsch *et al.* 2009, WO 2005/10068; WO 2009/029266, WO 2008/119776; WO 2009/111881; WO 2006/127537; US 2003/0104419; US 2008/0145841; US 2006/0035244; US 7,598,052; US 2008/0274457; US 2008/0213805). Among others, *Cyclin D2 (CCND2)*, *protein convertase 2 (PCSK2)*, and *prostate differentiation factor (PLAB)* have been reported to differentiate between FC and FA (Weber *et al.* 2005). With few exceptions (Weber *et al.* 2005; Fryknas *et al.* 2006; Fijarewicz *et al.* 2007; Prasad *et al.* 2008), most studies have relied on unsupervised methods, such as hierarchical clustering of list of differentially expressed genes, which are not appropriate methods to provide signatures that are robust across geographical locations or platforms that may affect the accuracy of the predictions made by a particular classifier (Simon 2006).
- Diagnosis of thyroid nodules to date may be performed using one or – more often - a combination of the below:
- Blood sample
 - Scintillation counting using a tracer to measure ionizing radiation
 - Ultrasound
 - Ultrasound guided biopsy
 - Cytology
 - Assessment of risk factors
 - Surgical removal of all or part of the thyroid gland (thyroidectomy)
- The high prevalence of thyroid nodules in Denmark and around the world leads to a high diagnostic activity, although there is no general consensus in the area on the most sensitive and specific diagnostic tool. In Denmark, more than 1500 thyroidectomies are performed annually, most due to nodular goiter and suspicion of neoplasia. Totally, 120-140 incidents of thyroid cancer are diagnosed annually in Denmark. This means

that the majority of thyroidectomies are performed in excess, with both economical and personal costs.

5 The many superfluous thyroidectomies are performed mainly due to non-conclusive biopsies or a finding of follicular neoplasia. Follicular neoplasia may prove to be either malignant (follicular thyroid carcinoma, FTC) or benign (follicular thyroid adenoma, FTA). Only the malignant subtype requires surgery, whereby an improved diagnostic answer from biopsies can help reduce the number of excess thyroidectomies.

10 Efforts to improve the pre-operative diagnosis of thyroid nodules are needed, in order to more efficiently distinguish benign from malignant nodules without the need for diagnostic surgery.

15 The present invention discloses a sensitive and specific means of distinction between follicular thyroid neoplasia subtypes, comprising follicular thyroid adenomas (benign), fetal adenomas (FeA), and follicular thyroid carcinomas (malignant). The inventors have found that a subset of specific mRNAs are differentially expressed in and associated with these subtypes of follicular thyroid neoplasia, efficiently separating the benign and the malignant subtypes of follicular thyroid neoplasia by employing mRNA
20 classifiers capable of predicting which of the above categories or classes a certain sample obtained from an individual belongs to.

The present invention is thus directed to the development of mRNA classifiers; A) that distinguish benign FTA from malignant FTC; and/or B) that distinguishes benign FTA
25 from malignant and pre-malignant FTC and fetal adenomas; and/or C) that distinguishes benign FTA, malignant FTC and pre-malignant FeA.

The terms distinction, differentiation, classification or characterisation of a sample is used herein as being capable of predicting with a high sensitivity and specificity if a
30 given sample of unknown diagnosis belongs to the class of benign FTA or malignant FTC, or belongs to the class of benign FTA or malignant and pre-malignant FTC and FeA. The output is given as a probability of belonging to either class of between 0-1.

35 The use of the herein disclosed mRNA classifiers may alleviate the need for the high number of diagnostic thyroidectomies performed on suspicion of all follicular neoplasias

including the benign adenomas, and is as such useful as a stand-alone or an 'add-on' method to the existing diagnostic methods currently used for characterising thyroid nodules. Further, an early diagnosis of a malignant condition may expedite treatment of patients presenting with a malignant nodule, i.e. placing this group of patients first in line for surgery.

Summary of invention

Thyroid nodules are frequent in the adult population. Efforts to improve the pre-operative diagnosis of thyroid nodules are needed, in order to more efficiently distinguish benign from malignant nodules without the need for diagnostic surgery.

The expression of mRNA is often deregulated in malignant cells and shows a highly tissue-specific pattern. A classifier based on an mRNA expression profile or signature, may be an ideal diagnostic tool to differentiate the malignant from the benign thyroid tumors.

The aim of the present invention is to develop two-way mRNA classifiers, which can accurately differentiate between subtypes of follicular thyroid neoplasms; the class of thyroid follicular adenomas (FTA) from the class of thyroid follicular carcinomas (FTC); and the class of thyroid follicular carcinomas (FTC) from the class of thyroid follicular adenomas (FTA) merged with fetal adenomas (FeA).

There is provided herein a system for the identification of a malignancy-specific signature of mRNAs that are differentially expressed relative to adenoma cells.

The present invention concerns molecular classifiers that can identify thyroid follicular carcinomas with high accuracy and specificity. The classifiers of the present invention are working on follicular nodules originating from different geographical locations and platforms. The use of the classifiers of the invention can improve pre-operative diagnosis.

The mRNA classifiers disclosed herein 1) distinguishes benign FTA from malignant FTC, 2) distinguishes benign FTA from malignant FTC and pre-malignant FeA, 3) distinguishes malignant FTC from merged benign FTA and pre-malignant FeA.

The above-mentioned mRNA classifiers have extraordinary performance as compared to contemporary standards; classifier 1 has an unprecedented sensitivity of 95% and specificity of 95%, while classifier 2 has a sensitivity of 89% and a specificity of 91%, and classifier 3 has a sensitivity of 89% and a specificity of 91%.

5

The classifiers in one embodiment comprise or consist of six or more mRNAs selected from the groups disclosed in tables 19, 20 and 21.

The classifiers in one embodiment comprise or consist of six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.

20

The mRNA classifiers may be applied *ex vivo* to a sample from a thyroid nodule of a human being, in order to improve the pre-operative diagnostic prognosis. This would reduce the current large number of diagnostic thyroid operations performed and expedite the necessary operations (i.e. on malignant nodules).

25

Accordingly, provided herein are methods for diagnosing whether a subject has, or is at risk of developing, follicular thyroid carcinoma and/or fetal adenoma, comprising the steps of extracting RNA from a sample collected from the thyroid of an individual and analysing the mRNA expression profile or signature of said sample, comprising of six or more mRNA sequences selected from the groups disclosed in tables 19, 20, and 21 or selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL,

35

ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1,
 5 DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.

Also provided are methods for determining the need for thyroidectomy in an individual presenting with a thyroid nodule by employing the mRNA classifiers disclosed herein.

10 The present invention is also directed to a device for measuring the expression level of six or more mRNAs, comprising or consisting of probes for a mRNA selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2,
 15 RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM,
 20 CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said device is used for classifying a sample obtained from a thyroid nodule of an individual.

25 Also provided is a system for performing a diagnosis on an individual with a thyroid nodule, comprising means for analysing the mRNA expression profile of the thyroid nodule, comprising six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR,
 30 CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9,
 35 MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A,

NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and means for determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma.

The present invention is also directed to a computer program product having a computer readable medium, said computer program product providing a system for predicting the diagnosis of an individual with a thyroid nodule, said computer program product comprising means for carrying out any of the steps of any of the methods as disclosed herein.

Definitions

Statistical classification is a procedure in which individual items are placed into groups based on quantitative information on one or more characteristics inherent in the items (referred to as traits, variables, characters, etc) and based on a training set of previously labeled items.

A classifier is a prediction model which may distinguish between or characterize samples by classifying a given sample into a predetermined class based on certain characteristics of said sample. A two-way classifier classifies a given sample into one of two predetermined classes, and a three-way classifier classifies a given sample into one of three predetermined classes.

'Collection media' as used herein denotes any solution suitable for collecting, storing or extracting of a sample for immediate or later retrieval of RNA from said sample.

'Deregulated' means that the expression of a gene or a gene product is altered from its normal baseline levels; comprising both up- and down-regulated.

Goiter: A swelling in the neck (just below the Adam's apple or larynx) due to an enlarged thyroid gland. Also denoted goitre (British), struma (Latin), or a bronchocele.

The term "Individual" refers to vertebrates, particular members of the mammalian species, preferably primates including humans. As used herein, 'subject' and 'individual' may be used interchangeably.

The term “Kit of parts” as used herein provides a device for measuring the expression level of six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and at least one additional component. The additional component may in one embodiment be means for extracting RNA, such as mRNA, from a sample; reagents for performing microarray analysis, reagents for performing QPCR analysis and/or instructions for use of the device and/or additional components.

The term “natural nucleotide” or “nucleotide” refers to any of the four deoxyribonucleotides, dA, dG, dT, and dC (constituents of DNA), and the four ribonucleotides, A, G, U, and C (constituents of RNA). Each natural nucleotide comprises or essentially consists of a sugar moiety (ribose or deoxyribose), a phosphate moiety, and a natural/standard base moiety. Natural nucleotides bind to complementary nucleotides according to well-known rules of base pairing (Watson and Crick), where adenine (A) pairs with thymine (T) or uracil (U); and where guanine (G) pairs with cytosine (C), wherein corresponding base-pairs are part of complementary, anti-parallel nucleotide strands. The base pairing results in a specific hybridization between predetermined and complementary nucleotides. The base pairing is the basis by which enzymes are able to catalyze the synthesis of an oligonucleotide complementary to the template oligonucleotide. In this synthesis, building blocks (normally the triphosphates of ribo or deoxyribo derivatives of A, T, U, C, or G) are directed by a template oligonucleotide to form a complementary oligonucleotide with the correct, complementary sequence. The recognition of an oligonucleotide sequence by its complementary sequence is mediated by corresponding and interacting bases forming base pairs. In nature, the specific interactions leading to base pairing are governed by the size of the bases and the pattern of hydrogen bond donors and acceptors of the bases. A large purine base (A or G) pairs with a small pyrimidine base

(T, U or C). Additionally, base pair recognition between bases is influenced by hydrogen bonds formed between the bases. In the geometry of the Watson-Crick base pair, a six membered ring (a pyrimidine in natural oligonucleotides) is juxtaposed to a ring system composed of a fused, six membered ring and a five membered ring (a purine in natural oligonucleotides), with a middle hydrogen bond linking two ring atoms, and hydrogen bonds on either side joining functional groups appended to each of the rings, with donor groups paired with acceptor groups.

As used herein, "nucleic acid" or "nucleic acid molecule" refers to polynucleotides, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), oligonucleotides, fragments generated by the polymerase chain reaction (PCR), and fragments generated by any of ligation, scission, endonuclease action, and exonuclease action. Nucleic acid molecules can be composed of monomers that are naturally-occurring nucleotides (such as DNA and RNA), or analogs of naturally-occurring nucleotides (e.g. alpha-enantiomeric forms of naturally-occurring nucleotides), or a combination of both.

Modified nucleotides can have alterations in sugar moieties and/or in pyrimidine or purine base moieties. Sugar modifications include, for example, replacement of one or more hydroxyl groups with halogens, alkyl groups, amines, and azido groups, or sugars can be functionalized as ethers or esters. Moreover, the entire sugar moiety can be replaced with sterically and electronically similar structures, such as aza-sugars and carbocyclic sugar analogs. Examples of modifications in a base moiety include alkylated purines and pyrimidines, acylated purines or pyrimidines, or other well-known heterocyclic substitutes. Nucleic acid monomers can be linked by phosphodiester bonds or analogs of such linkages. Analogues of phosphodiester linkages include phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate, and the like. The term "nucleic acid molecule" also includes e.g. so-called "peptide nucleic acids," which comprise naturally-occurring or modified nucleic acid bases attached to a polyamide backbone. Nucleic acids can be either single stranded or double stranded. A "polypeptide" or "protein" is a polymer of amino acid residues preferably joined exclusively by peptide bonds, whether produced naturally or synthetically. The term "polypeptide" as used herein covers proteins, peptides and polypeptides, wherein said proteins, peptides or polypeptides may or may not have been post-translationally modified. Post-translational modification may for example be phosphorylation, methylation and glycosylation.

Thyroidectomy: A thyroidectomy involves the surgical removal of all or part of the thyroid gland.

A 'probe' as used herein refers to a hybridization probe. A hybridization probe is a (single-stranded) fragment of DNA or RNA of variable length (usually 100-1000 bases long), which is used in DNA or RNA samples to detect the presence of nucleotide sequences (the DNA target) that are complementary to the sequence in the probe. The probe thereby hybridizes to single-stranded nucleic acid (DNA or RNA) whose base sequence allows probe-target base pairing due to complementarity between the probe and target. To detect hybridization of the probe to its target sequence, the probe is tagged (or labelled) with a molecular marker of either radioactive or fluorescent molecules. DNA sequences or RNA transcripts that have moderate to high sequence similarity to the probe are then detected by visualizing the hybridized probe.

Hybridization probes used in DNA microarrays refer to DNA covalently attached to an inert surface, such as coated glass slides or gene chips, and to which a mobile cDNA target is hybridized.

Due to the imprecision of standard analytical methods, molecular weights and lengths of polymers are understood to be approximate values. When such a value is expressed as "about" X or "approximately" X, the stated value of X will be understood to be accurate to +/- 20%, such as +/- 10%, for example +/- 5%.

Follicular thyroid carcinoma (FTC) and follicular carcinoma (FC) and thyroid follicular carcinoma are used interchangeably herein; follicular thyroid adenoma (FTA) and follicular adenoma (FA) and thyroid follicular adenoma are used interchangeably herein; and fetal adenomas (FeA) and microfollicular adenomas are used interchangeably herein.

Detailed description of the invention

The thyroid gland

The thyroid is one of the largest endocrine glands in the body. This gland is found in the neck inferior to the thyroid cartilage ('Adam's apple' in men) and at approximately the same level as the cricoid cartilage. The thyroid controls how quickly the body burns energy, makes proteins, and how sensitive the body should be to other hormones.

The thyroid participates in these processes by producing thyroid hormones, principally thyroxine (T_4) and triiodothyronine (T_3). These hormones regulate the rate of

metabolism and affect the growth and rate of function of many other systems in the body. Iodine is an essential component of both T_3 and T_4 . The thyroid also produces the hormone calcitonin, which plays a role in calcium homeostasis. The thyroid is in turn controlled by the hypothalamus and pituitary.

5

The thyroid is composed of spherical follicles that selectively absorb iodine (as iodide ions, I^-) from the blood for production of thyroid hormones. Twenty-five percent of all the body's iodide ions are in the thyroid gland. Inside the follicles, colloids rich in a protein called thyroglobulin serve as a reservoir of materials for thyroid hormone production and, to a lesser extent, act as a reservoir for the hormones themselves. The follicles are surrounded by a single layer of thyroid epithelial cells (or 'follicular cells'), which secrete T_3 and T_4 . When the gland is not secreting T_3/T_4 (inactive), the epithelial cells range from low columnar to cuboidal cells. When active, the epithelial cells become tall columnar cells. Scattered among follicular cells and in spaces between the spherical follicles are another type of thyroid cell, parafollicular cells or C cells, which secrete calcitonin.

Thyroxine (T_4) is synthesised by the follicular cells from free tyrosine and on the tyrosine residues of the protein called thyroglobulin (TG). Iodine is captured with the "iodine trap" by the hydrogen peroxide generated by the enzyme thyroid peroxidase (TPO) and linked to the 3' and 5' sites of the benzene ring of the tyrosine residues on TG, and on free tyrosine. Upon stimulation by the thyroid-stimulating hormone (TSH), the follicular cells reabsorb TG and proteolytically cleave the iodinated tyrosines from TG, forming T_4 and T_3 (in T_3 , one iodine is absent compared to T_4), and releasing them into the blood. Deiodinase enzymes convert T_4 to T_3 . Thyroid hormone that is secreted from the gland is about 90% T_4 and about 10% T_3 .

Cells of the brain are a major target for the thyroid hormones T_3 and T_4 . Thyroid hormones play a particularly crucial role in brain maturation during fetal development. A transport protein (OATP1C1) has been identified that seems to be important for T_4 transport across the blood brain barrier. A second transport protein (MCT8) is important for T_3 transport across brain cell membranes.

In the blood, T_4 and T_3 are partially bound to thyroxine-binding globulin, transthyretin and albumin. Only a very small fraction of the circulating hormone is free (unbound) -

T4 0.03% and T3 0.3%. Only the free fraction has hormonal activity. As with the steroid hormones and retinoic acid, thyroid hormones cross the cell membrane and bind to intracellular receptors (α_1 , α_2 , β_1 and β_2), which act alone, in pairs or together with the retinoid X-receptor as transcription factors to modulate DNA transcription. Up to 80% of the T4 is converted to T3 by peripheral organs such as the liver, kidney and spleen. T3 is about ten times more active than T4.

The production of thyroxine and triiodothyronine is regulated by thyroid-stimulating hormone (TSH), released by the anterior pituitary (that is in turn released as a result of TRH release by the hypothalamus). The thyroid and thyrotropes form a negative feedback loop: TSH production is suppressed when the T4 levels are high, and vice versa. The TSH production itself is modulated by thyrotropin-releasing hormone (TRH), which is produced by the hypothalamus and secreted at an increased rate in situations such as cold (in which an accelerated metabolism would generate more heat). TSH production is blunted by somatostatin (SRIH), rising levels of glucocorticoids and sex hormones (estrogen and testosterone), and excessively high blood iodide concentration.

An additional hormone produced by the thyroid contributes to the regulation of blood calcium levels. Parafollicular cells produce calcitonin in response to hypercalcemia. Calcitonin stimulates movement of calcium into bone, in opposition to the effects of parathyroid hormone (PTH). However, calcitonin seems far less essential than PTH, as calcium metabolism remains clinically normal after removal of the thyroid, but not the parathyroids.

25

Thyroid nodule

Thyroid nodules are lumps which commonly arise within an otherwise normal thyroid gland. Often these abnormal growths of thyroid tissue are located at the edge of the thyroid gland so they can be felt as a lump in the throat. When they are large or when they occur in very thin individuals, they may even be seen as a lump in the front of the neck. Thyroid nodules are extremely common and almost 50% of people have had one, but they are usually only detected by a general practitioner during the course of a health examination, or through a different affliction. Only a small percentage of lumps in the neck are malignant (less than 1%), and most thyroid nodules are benign.

35

Thyroid neoplasia

Neoplasia is the abnormal proliferation of cells, resulting in a structure known as a neoplasm. The growth of this clone of cells exceeds, and is uncoordinated with, that of the normal tissues around it. It usually causes a lump or tumor. Neoplasms may be

5 benign, pre-malignant or malignant.

Thyroid neoplasia may be benign (adenoma) or malignant (carcinoma), with only the malignant requiring surgery.

10 *Benign neoplasia*

A thyroid adenoma, or solitary thyroid nodule, is a benign tumor of the thyroid gland. A thyroid adenoma is distinguished from a multinodular goiter of the thyroid in that an adenoma is typically solitary, and is a neoplasm resulting from a genetic mutation (or other genetic abnormality) in a single precursor cell. In contrast, a multinodular goiter is

15 usually thought to result from a hyperplastic response of the entire thyroid gland to a stimulus, such as iodine deficiency. A thyroid adenoma may be clinically silent, or it may be a functional tumor, producing excessive thyroid hormone. In this case, it may result in symptomatic hyperthyroidism, and may be referred to as a toxic thyroid adenoma. Careful pathological examination may be necessary to distinguish a thyroid

20 adenoma from a minimally invasive follicular thyroid carcinoma.

Malignant neoplasia

Thyroid cancer is more frequent in females at a ratio of three to one. Thyroid cancer can occur in any age group, although it is most common after age 30 and its

25 aggressiveness increases significantly in older patients. The majority of patients present with a nodule on their thyroid which typically does not cause symptoms. When a thyroid cancer begins to grow within a thyroid gland, it almost always does so within a discrete nodule within the thyroid. Scintigraphically cold nodules are more likely to be cancerous, however only a small part of the cold nodules are diagnosed as cancer.

30

Thyroid cancer or carcinoma refers to any of four kinds of malignant tumors of the thyroid gland: papillary, follicular, medullary or anaplastic. Papillary and follicular tumors are the most common. They grow slowly and may recur, but are generally not fatal in patients under 45 years of age. Medullary tumors have a good prognosis if

restricted to the thyroid gland and a poorer prognosis if metastasis occurs. Anaplastic tumors are fast-growing and respond poorly to therapy.

In Denmark, 120-140 incidents of thyroid cancer are diagnosed each year. Of these, 63% are papillary carcinomas, 18% are follicular neoplasia, 7% are medullary neoplasia, 8% are anaplastic (undifferentiated); leaving 4% designated as 'others' (including metastasis, lymphoma, squamous cell carcinoma, sarcoma). The follicular and papillary types together can be classified as "differentiated thyroid cancer". These types have a more favorable prognosis than the medullary and undifferentiated types.

Papillary thyroid carcinoma

Papillary thyroid cancer is generally the most common type of thyroid cancer. It occurs more frequently in women and presents in the 30-40 year age group. It is also the predominant cancer type in children with thyroid cancer, and in patients with thyroid cancer who have had previous radiation to the head and neck. Papillary microcarcinoma is a subset of papillary thyroid cancer defined as measuring less than or equal to 1cm. Papillary thyroid carcinoma commonly metastasizes to cervical lymph nodes.

Thyroglobulin can be used as a tumor marker for well-differentiated papillary thyroid cancer. HBME-1 (human mesothelial cell 1) staining may be useful for differentiating papillary carcinomas from follicular carcinomas; in papillary lesions it tends to be positive.

Surgical treatment includes either hemithyroidectomy (or unilateral lobectomy) or isthmectomy (removing the band of tissue (or isthmus) connecting the two lobes of the thyroid), which is sometimes indicated with minimal disease (diameter up to 1.0 centimeters). For gross disease (diameter over 1 centimeter), total thyroidectomy, and central compartment lymph node removal is the therapy of choice. As papillary carcinoma is a multifocal disease, hemithyroidectomy may leave disease in the other lobe and total thyroidectomy reduces the risk of recurrence.

Follicular thyroid carcinoma

Follicular thyroid cancer is a form of thyroid cancer which occurs more commonly in women of over 50 years. Follicular carcinoma is considered more malignant

(aggressive) than papillary carcinoma. It occurs in a slightly older age group than papillary cancer and is also less common in children. In contrast to papillary cancer, it occurs only rarely after radiation therapy. Mortality is related to the degree of vascular invasion. Age is a very important factor in terms of prognosis. Patients over 40 have a more aggressive disease and typically the tumor does not concentrate iodine as well as in younger patients. Vascular invasion is characteristic for follicular carcinoma and therefore distant metastasis is more common. Lung, bone, brain, liver, bladder, and skin are potential sites of distant spread. Lymph node involvement is far less common than in papillary carcinoma.

Unlike papillary thyroid cancer, follicular thyroid cancer is today difficult to diagnose without performing surgery because there are no characteristic changes in the way the thyroid cells look; i.e. it is not possible to accurately distinguish between follicular thyroid adenoma and carcinoma on cytological grounds. Rather, the only way to tell if a follicular cell nodule (or neoplasm) is cancer, is to look at the entire capsule around the nodule and see if there is any sign of invasion. A fine needle aspiration (FNA) biopsy cannot at present distinguish cytologically between follicular adenoma, follicular carcinoma and a completely benign condition called nontoxic nodular goiter. Even a coarse needle biopsy, which is typically more accurate than a FNA, cannot always provide an answer since it is only able to differentiate between a follicular neoplasm (which includes both adenoma and carcinoma) versus nontoxic nodular goiter about 40% of the time. These biopsies can only look at individual cells and not the entire capsule. If fine needle aspiration cytology suggests follicular neoplasm, thyroid lobectomy is today performed to establish the histopathological diagnosis. This difficulty in diagnosis is one of the most frustrating areas for physicians who study thyroid disease today, because it means that surgery is most often the only way of definitively diagnosing a thyroid nodule.

It is an object of the present invention to disclose a method for more efficiently distinguishing between the malignant and benign subtypes of follicular neoplasm; thus improving the pre-operative diagnosis of this condition and reducing the number of diagnostic surgeries required. This is achieved by providing specific mRNA classifiers that may distinguish between the benign follicular adenomas and the malignant follicular carcinomas; and/or distinguish between the benign follicular adenomas and the malignant follicular carcinomas and the fetal adenomas; and/or distinguish between

the benign follicular adenomas and the combined group of malignant follicular carcinomas and pre-malignant or fetal adenomas (or microfollicular adenomas).

5 Treatment is usually surgical, followed by radioiodine. Unilateral hemithyroidectomy (removal of one entire lobe of the thyroid) is uncommon due to the aggressive nature of this form of thyroid cancer, but may be indicated for achieving the diagnosis. Total thyroidectomy is almost automatic with this diagnosis. This is invariably followed by radioiodine treatment following two weeks of a low iodine diet. Occasionally treatment must be repeated if annual scans indicate remaining cancerous tissue. Minimally
10 invasive thyroidectomy has been used in recent years in cases where the nodules are small.

Fetal adenoma (microfollicular adenomas or follicular fetal adenoma) is a subgroup of follicular neoplasms with a potential to transform into malignancy. The term 'fetal
15 adenoma' was coined to designate certain nodular tumors of the thyroid gland, which was originally believed to arise from fetal cell rests. With an advance in knowledge, however, the concept of a fetal origin for these nodules has largely been discarded. Today it has come to designate a distinctive type of nodule, on the general features of which most observers are agreed. They begin as masses of thyroid tissue which has
20 never reached an adult stage.

Fetal adenoma represents a distinct histopathological entity. Their malignant potential is poorly characterized, but since they exhibit a high degree (58%) of aneuploidy, they may progress to malignancy. In agreement with this assumption it is known that about
25 5 percent of fetal adenomas prove to be follicular cancers with careful, histopathological study.

Hurthle cell thyroid cancer is often considered a variant of follicular cell carcinoma. Hurthle cell forms are more likely than follicular carcinomas to be bilateral and
30 multifocal and to metastasize to lymph nodes. Like follicular carcinoma, unilateral hemithyroidectomy is performed for non-invasive disease, and total thyroidectomy for invasive disease.

Medullary thyroid carcinoma

Medullary thyroid cancer (MTC) is a form of thyroid carcinoma which originates from the parafollicular cells (C cells), which produce the hormone calcitonin. Approximately 25% the cancer develops in families. When MTC occurs by itself it is termed familial MTC; when it coexists with tumors of the parathyroid gland and medullary component of the adrenal glands (pheochromocytoma) it is called multiple endocrine neoplasia type 2 (MEN2).

While the increased serum concentration of calcitonin is not harmful, it is useful as a marker which can be tested in blood. A second marker, carcinoembryonic antigen (CEA), also produced by medullary thyroid carcinoma, is released into the blood and it is useful as a serum or blood tumor marker. In general measurement of serum CEA is less sensitive than serum calcitonin for detecting the presence of a tumor, but has less minute to minute variability and is therefore useful as an indicator of tumor mass.

Mutations in the RET proto-oncogene ("rearranged during transfection"), located on chromosome 10, lead to the expression of a mutated receptor tyrosine kinase protein, termed RET. RET is involved in the regulation of cell growth and development and its mutation is responsible for nearly all cases of hereditary or familial medullary thyroid carcinoma.

Surgery can be effective when the condition is detected early, but a risk for recurrence remains. Unlike differentiated thyroid carcinoma, there is no role for radioiodine treatment in medullary-type disease. External beam radiotherapy should be considered for patients at high risk of regional recurrence, even after optimum surgical treatment. Also, clinical trials of several new tyrosine kinase inhibitors are now being studied.

The prognosis of MTC is poorer than that of follicular and papillary thyroid cancer when it has metastasized (spread) beyond the thyroid gland.

Anaplastic thyroid carcinoma

Anaplastic thyroid cancer (ATC) or undifferentiated thyroid cancer is a form of thyroid cancer which has a very poor prognosis due to its aggressive behavior and resistance to cancer treatments. It rapidly invades surrounding tissues (such as the trachea). The presence of regional lymphadenopathy in older patients in whom a characteristic

vesicular appearance of the nuclei is revealed would support a diagnosis of anaplastic carcinoma.

5 The median survival from diagnosis ranges from 3 to 7 months, with worse prognosis associated with large tumours, distant metastases, acute obstructive symptoms, and leucocytosis. In the 18–24% of patients whose tumour seems both confined to the neck and grossly resectable, complete surgical resection followed by adjuvant radiotherapy and chemotherapy could yield a 75–80% survival at 2 years.

10 Unlike its differentiated counterparts, anaplastic thyroid cancer is highly unlikely to be curable either by surgery or by any other treatment modality, and is in fact usually unresectable due to its high propensity for invading surrounding tissues. Palliative treatment consists of radiation therapy usually combined with chemotherapy. New drugs are in clinical trials that may improve chemotherapy treatment.

15

Diagnosing thyroid neoplasia at present

Most often the first symptom of thyroid cancer is a nodule in the thyroid region of the neck. However, many adults have small undetected nodules in their thyroids. Typically fewer than 5% of these nodules are found to be malignant. Sometimes the first sign is an enlarged lymph node. Later possible symptoms are pain in the anterior region of the neck and changes in voice. Thyroid cancer is usually found in a euthyroid patient (having normal thyroid function), but symptoms of hyperthyroidism may be associated with a large or metastatic well-differentiated tumor.

20

25 Diagnosing of thyroid nodules to date may be performed using one or – more often - a combination of the below diagnostic methods:

- *Scintillation counting* using a tracer to measure ionizing radiation (using technetium Tc or ionizing Iodine I¹³¹ or I¹²³). 85% of nodules will be scintigraphically 'cold'; i.e. not accumulating the tracer. Of these, 5% will be malignant. Hot nodules are signs of non-cancerous nodules.
- *Blood sample*. Measurement of thyroid stimulating hormone (TSH) and anti-thyroid antibodies will help decide if there is a functional (non-cancerous) thyroid disease present. The possibility of a nodule which secretes thyroid hormone (which is less likely to be cancer) or hypothyroidism is investigated by measuring thyroid stimulating hormone (TSH), and the thyroid hormones

30

35

thyroxine (T4) and triiodothyronine (T3). Tests for serum thyroid auto-antibodies are sometimes done as these may indicate autoimmune thyroid disease (which can mimic nodular disease).

- 5 - *Ultrasound imaging.* Features that may be distinguished using ultrasound relies on an assessment from the operator, and includes relating a feature with a probability (rare to high) of malignancy. Features include amongst others lymphadenopathies, invasion on adjacent structure, poorly defined margins, cystic nodule, blood flow level and microcalcifications.
- 10 - *Cytology/histology* of resected thyroid nodule (e.g. thyroidectomy or biopsy).
- 10 - *Assessment of risk factors*, comprising the occurrence of thyroid cancer in the family, age below 20 or above 70 years, male gender, previous irradiation to the neck and/or head area, large nodule (>4 cm), fast growing nodule, firm or hard texture, fixation to surrounding structures, compression symptoms (hoarse voice, dysphagia, dyspnea) and regional lymphadenopathy.

15

While the above diagnostic tools may render probable that a nodule is indeed cancerous, it is not straight forward to distinguish between the four kinds of malignant tumors of the thyroid gland (papillary, follicular, medullary or anaplastic), and further to diagnose malignant follicular thyroid cancer without performing surgery, because it is at present not possible to accurately distinguish between follicular thyroid adenoma and carcinoma on cytological grounds. Indeed, diagnostic surgery is the only certain way to establish a correct diagnosis on a thyroid nodule.

20

The method disclosed herein provides a tool for improving the pre-operative diagnosis of thyroid nodules, in particular thyroid follicular neoplasm, thus reducing the number of diagnostic surgeries required. Specific mRNA classifiers are provided that may distinguish between the benign follicular adenomas and the malignant follicular carcinomas; and/or distinguish between the benign follicular adenomas and the malignant follicular carcinomas and the pre-malignant fetal adenomas and/or distinguish between the benign follicular carcinomas and the merged group of malignant follicular adenomas and pre-malignant fetal adenomas (or microfollicular adenoma).

25

30

The mRNA classifiers as disclosed herein may in one embodiment be used in the clinic alone.

In another embodiment, the mRNA classifiers as disclosed herein may be used in the clinic as an add-on or supplementary diagnostic tool or method, which improves the pre-operative diagnosis of thyroid nodules by combining the output of said mRNA classifier with the output of one or more of the above-mentioned conventional diagnostic techniques to improve the accuracy of said pre-operative diagnosis of thyroid neoplasms.

Nucleic Acids

A nucleic acid is a biopolymeric macromolecule composed of chains of monomeric nucleotides. In biochemistry these molecules carry genetic information or form structures within cells. The most common nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Each nucleotide consists of three components: a nitrogenous heterocyclic base (the nucleobase component), which is either a purine or a pyrimidine; a pentose sugar (backbone residues); and a phosphate group (internucleoside linkers). A nucleoside consists of a nucleobase (often simply referred to as a base) and a sugar residue in the absence of a phosphate linker. Nucleic acid types differ in the structure of the sugar in their nucleotides - DNA contains 2-deoxyribose while RNA contains ribose (where the only difference is the presence of a hydroxyl group). Also, the nitrogenous bases found in the two nucleic acid types are different: adenine, cytosine, and guanine are found in both RNA and DNA, while thymine only occurs in DNA and uracil only occurs in RNA. Other rare nucleic acid bases can occur, for example inosine in strands of mature transfer RNA. Nucleobases are complementary, and when forming base pairs, must always join accordingly: cytosine-guanine, adenine-thymine (adenine-uracil when RNA). The strength of the interaction between cytosine and guanine is stronger than between adenine and thymine because the former pair has three hydrogen bonds joining them while the latter pair has only two. Thus, the higher the GC content of double-stranded DNA, the more stable the molecule and the higher the melting temperature.

Nucleic acids are usually either single-stranded or double-stranded, though structures with three or more strands can form. A double-stranded nucleic acid consists of two single-stranded nucleic acids held together by hydrogen bonds, such as in the DNA double helix. In contrast, RNA is usually single-stranded, but any given strand may fold

back upon itself to form secondary structure as in tRNA and rRNA. Messenger ribonucleic acid (mRNA) is transcribed from a DNA template and carries the coding information for protein synthesis.

- 5 The sugars and phosphates in nucleic acids are connected to each other in an alternating chain, linked by shared oxygens, forming a phosphodiester bond. In conventional nomenclature, the carbons to which the phosphate groups attach are the 3' end and the 5' end carbons of the sugar. This gives nucleic acids polarity. The bases extend from a glycosidic linkage to the 1' carbon of the pentose sugar ring. Bases are
10 joined through N-1 of pyrimidines and N-9 of purines to 1' carbon of ribose through N- β glycosyl bond.

Classifier

- 15 Classifiers are relationships between sets of input variables, usually known as features, and discrete output variables, known as classes. Classes are often centered on the key questions of who, what, where and when. A classifier can intuitively be thought of as offering an opinion about whether, for instance, an individual associated with a given feature set is a member of a given class.

- 20 In other words, a classifier is a predictive model that attempts to describe one column (the label) in terms of others (the attributes). A classifier is constructed from data where the label is known, and may be later applied to predict label values for new data where the label is unknown. Internally, a classifier is an algorithm or mathematical formula that predicts one discrete value for each input row. For example, a classifier built from
25 a dataset of iris flowers could predict the type of a presented iris given the length and width of its petals and stamen. Classifiers may also produce probability estimates for each value of the label. For example, a classifier built from a dataset of cars could predict the probability that a specific car was built in the United States.

30 *Sensitivity and specificity*

- Sensitivity and specificity are statistical measures of the performance of a binary classification test. The sensitivity (also called recall rate in some fields) measures the proportion of actual positives which are correctly identified as such (i.e. the percentage of sick people who are identified as having the condition); and the specificity measures
35 the proportion of negatives which are correctly identified (i.e. the percentage of well

people who are identified as not having the condition). They are closely related to the concepts of type I and type II errors.

5 For any test, there is usually a trade-off between each measure. For example in a manufacturing setting in which one is testing for faults, one may be willing to risk discarding functioning components (low specificity), in order to increase the chance of identifying nearly all faulty components (high sensitivity). This trade-off can be represented graphically using a ROC curve.

$$10 \text{ sensitivity} = \frac{\text{number of True Positives}}{\text{number of True Positives} + \text{number of False Negatives}}$$

A sensitivity of 100% means that the test recognizes all sick people as such. Thus in a high sensitivity test, a negative result is used to rule out the disease.

15 Sensitivity alone does not tell us how well the test predicts other classes (that is, about the negative cases). In the binary classification, as illustrated above, this is the corresponding specificity test, or equivalently, the sensitivity for the other classes. Sensitivity is not the same as the positive predictive value (ratio of true positives to combined true and false positives), which is as much a statement about the proportion of actual positives in the population being tested as it is about the test.

20 The calculation of sensitivity does not take into account indeterminate test results. If a test cannot be repeated, the options are to exclude indeterminate samples from analyses (but the number of exclusions should be stated when quoting sensitivity), or, alternatively, indeterminate samples can be treated as false negatives (which gives the worst-case value for sensitivity and may therefore underestimate it).

$$\text{specificity} = \frac{\text{number of True Negatives}}{\text{number of True Negatives} + \text{number of False Positives}}$$

30 A specificity of 100% means that the test recognizes all healthy people as healthy. Thus a positive result in a high specificity test is used to confirm the disease. The maximum is trivially achieved by a test that claims everybody healthy regardless of the true condition. Therefore, the specificity alone does not tell us how well the test

recognizes positive cases. We also need to know the sensitivity of the test to the class, or equivalently, the specificities to the other classes. A test with a high specificity has a low Type I error rate.

- 5 Specificity is sometimes confused with the precision or the positive predictive value, both of which refer to the fraction of returned positives that are true positives. The distinction is critical when the classes are different sizes. A test with very high specificity can have very low precision if there are far more true negatives than true positives, and vice versa.

10

mRNA classifiers according to the present invention

- The mRNA classifiers according to the present invention are the relationships between sets of input variables, i.e. the mRNA expression in a sample of an individual, and discrete output variables, i.e. distinction between a benign and malignant, or a benign, 15 a malignant and a pre-malignant, or a benign and malignant/pre-malignant condition of the thyroid. Thus, the classifier assigns a given sample to a given class with a given probability.

- Distinction, differentiation or characterisation of a sample is used herein as being 20 capable of predicting with a high sensitivity and specificity if a given sample of unknown diagnosis belongs to one of two classes (two-way classifier), or belongs to one of three classes (three-way classifier).

- In one embodiment, the mRNA classifier is capable of predicting with a high sensitivity 25 and specificity if a given sample of unknown diagnosis belongs to 1) the class of benign FTA or malignant FTC; or 2) belongs to the class of benign FTA or malignant FTC or pre-malignant FeA; or 3) belongs to the class of malignant FTC or merged benign FTA and pre-malignant FeA

- 30 There is provided herein a system for the identification of a malignancy-specific signature of mRNAs that are differentially expressed relative to adenoma cells.

- Platt's probabilistic outputs for Support Vector Machines (Platt, J. in Smola, A.J, et al. (eds.) Advances in large margin classifiers. Cambridge, 2000; incorporated herein by 35 reference) is useful for applications that require posterior class probabilities. Also

incorporated by reference herein is Platt J. Advances in Large Classifiers. Cambridge, MA: MIT Press, 1999.

5 The output of the mRNA classifiers is given as a probability of belonging to either class of between 0-1 (prediction probability). If the value for a sample is 0.5, no prediction is made. A number or value of between 0.51 to 1.0 for a given sample means that the sample is predicted to belong to the class in question, e.g. FTA; and the corresponding value of 0.0 to 0.49 for the second class in question, e.g. FTC means that the sample is predicted *not* to belong to the class in question.

10

In one embodiment, the prediction probabilities for a sample to belong to a certain class is a number falling in the range of from 0 to 1, such as from 0.0 to 0.1, for example 0.1 to 0.2, such as 0.2 to 0.3, for example 0.3 to 0.4, such as 0.4 to 0.49, for example 0.5, such as 0.51 to 0.6, for example 0.6 to 0.7, such as 0.7 to 0.8, for example 0.8 to 0.9, such as 0.9 to 1.0.

15

In one embodiment, the prediction probability for a sample to belong to the FTA class is a number falling in the range of from 0 to 0.49, 0.5 or from 0.51 to 1.0. In another embodiment, the prediction probability for a sample to belong to the FTC class is a number between from 0 to 0.49, 0.5 or between from 0.51 to 1.0. In yet another embodiment, the prediction probability for a sample to belong to the merged class of FTA and FeA is a number between from 0 to 0.49, 0.5 or from 0.51 to 1.0.

20

The classifiers according to the present invention may in one embodiment consist of 6 mRNAs, such as 7 mRNAs, for example 8 mRNAs, such as 9 mRNAs, for example 10 mRNAs.

25

In one aspect, the present invention relates to mRNA classifiers for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL,

30

35

ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1,
5 DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and distinguishes between the classes thyroid follicular adenoma and thyroid follicular carcinoma, wherein said distinction is given as a prediction probability for said sample of belonging to either class, said probability being a number falling in the range of from 0 to 1.

10

In another aspect, the present invention relates to mRNA classifiers for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the group consisting of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2,
15 CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and SLCO2A1, and distinguishes between the classes thyroid follicular adenoma and thyroid follicular carcinoma, wherein said distinction is given as a prediction probability for said sample of belonging to either class, said probability being a number falling in the range of from 0 to 1.

20

In one embodiment, said mRNA classifier comprises or consists of six or more mRNAs selected from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.

25

In another embodiment, said mRNA classifier comprises or consists of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.

30

In a further embodiment, the mRNA classifiers comprise one or more mRNAs selected from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.

The mRNA classifiers according to the present invention in one embodiment comprises less than 120 mRNAs, such as less than 110 mRNAs, for example less than 100 mRNAs, such as less than 90 mRNAs, for example less than 80 mRNAs, such as less than 70 mRNAs, for example less than 60 mRNAs, such as less than 50 mRNAs, for example less than 40 mRNAs, such as less than 30 mRNAs, for example less than 20 mRNAs, such as less than 10 mRNAs.

In one embodiment said mRNA classifier comprises or consists of between 6 to 10 mRNAs according to the present invention, for example 10 to 15, such as 15 to 20, for example 20 to 25, such as 25 to 30, for example 30 to 35, such as 35 to 40, for example 40 to 45, such as 45 to 50, for example 50 to 55, such as 55 to 60, for example 60 to 65, such as 65 to 70, for example 70 to 75, such as 75 to 80, for example 80 to 85, such as 85 to 90, for example 90 to 95, such as 95 to 100, for example 100 to 110, such as 110 to 120 mRNAs according to the present invention.

In a particular embodiment, said mRNA classifier consists of 64 mRNAs according to the present invention.

The expression of each mRNA in each thyroid sample used for constructing the mRNA classifiers as defined herein were determined, and the combined pattern of expression of the herein disclosed mRNAs forms the basis of the classifier model capable of predicting a diagnosis.

In an embodiment, an alteration of the expression profile or signature of one or more of the mRNAs of the mRNA classifier is associated with the sample being classified as thyroid follicular adenoma or thyroid follicular carcinoma; or as thyroid follicular carcinoma or thyroid follicular adenoma merged with fetal adenoma.

In one embodiment, the present invention relates to an mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1,

LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, 5 NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and distinguishes between thyroid follicular adenoma and thyroid follicular carcinoma.

10 In one embodiment, the mRNA classifier is indicative of thyroid follicular carcinoma in the event that ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, and/or UBE2C expression is up-regulated.

15 In one embodiment, the mRNA classifier is indicative of thyroid follicular carcinoma in the event that AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and/or SLCO2A1 expression is down-regulated.

20 In another embodiment, the present invention relates to an mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, 25 SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, 30 MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and distinguishes between thyroid follicular carcinoma and thyroid follicular adenoma merged with fetal adenoma.

The mRNA classifiers disclosed herein in a preferred embodiment has a sensitivity of at least 85%, such as at least 86%, for example at least 87%, such as at least 88%, for example at least 89%, such as at least 90%, for example at least 91%, such as at least 92%, for example at least 93%, such as at least 94%, for example at least 95%.

5

The mRNA classifiers disclosed herein in a preferred embodiment has a specificity of at least 85%, such as at least 86%, for example at least 87%, such as at least 88%, for example at least 89%, such as at least 90%, for example at least 91%, such as at least 92%, for example at least 93%, such as at least 94%, for example at least 95%.

10

In another embodiment, the mRNA classifier is capable of predicting if a given sample of unknown diagnosis belongs to the class of benign FTA, the class of malignant FTC, or the class of pre-malignant FeA.

15

The output of the mRNA classifier is given as a probability of belonging to either class of between 0-1 (prediction probability). If the number or value for a sample is 0.33, no prediction is made.

20

It is thus a further aspect of the present invention to provide a mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and distinguishes between the classes thyroid follicular adenoma, thyroid follicular carcinoma, and fetal adenoma, wherein said distinction is given as a prediction probability for said sample of belonging to either class, said probability being a number falling in the range of from 0 to 1.

35

The mRNA classifier according to the present invention preferably comprises less than 120 mRNAs, such as less than 110 mRNAs, for example less than 100 mRNAs, such as less than 90 mRNAs, for example less than 80 mRNAs, such as less than 70 mRNAs, for example less than 60 mRNAs, such as less than 50 mRNAs, for example less than 40 mRNAs, such as less than 30 mRNAs, for example less than 20 mRNAs, such as less than 10 mRNAs.

A model for predicting a diagnosis by employing the mRNA classifier of the present invention

In one aspect, the present invention relates to a model for predicting the diagnosis of an individual with a thyroid nodule, comprising

- i) providing a set of input data to the mRNA classifier according to the present invention, and
- ii) determining if said individual has a condition selected from the group of thyroid follicular adenoma, thyroid follicular carcinoma, and/or fetal adenoma.

In one embodiment, said input data comprises or consists of the mRNA expression profile of six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.

In another embodiment said input data comprises or consists of the mRNA expression profile of six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A,

ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, 5 EBAG9, MCFD2, PLDN, TCEAL4, and ZNF330.

In a further embodiment said input data comprises or consists of the mRNA expression profile of six or more mRNAs selected from the group consisting of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, 10 MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and SLCO2A1.

15 In yet another embodiment, the model according to the present invention comprises or consists of six or more mRNAs selected from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.

Methods employing the mRNA classifier of the present invention

20 In one aspect, the present invention relates to a method for determining the presence of a malignant and/or pre-malignant condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample selected from the groups disclosed in tables 19, 20, and 21 or the groups consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, 25 DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, 30 SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said expression level of said mRNAs is associated with thyroid follicular carcinoma or

thyroid follicular adenoma merged with fetal adenoma by predicting said association according to the mRNA classifier disclosed herein.

5 In another embodiment, the present invention relates to a method for determining the presence of a malignant condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, 10 SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, 15 MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and wherein said expression level of said mRNAs is associated with thyroid follicular carcinoma.

20

In yet another embodiment, the present invention relates to a method for determining the presence of a malignant or pre-malignant condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the 25 groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, 30 PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, 35 NR4A3, SLC26A4, and SLCO2A1.

In one aspect, the present invention relates to a method for determining the presence of a benign condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said expression level of said mRNAs is associated with thyroid follicular adenoma by predicting said association according to the mRNA classifier disclosed herein.

In another embodiment, the present invention relates to a method for determining the presence of a benign condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the groups disclosed in tables 19, 20, and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and wherein said expression level of said mRNAs is associated with thyroid follicular adenoma by predicting said association.

The invention in a further aspect relates to a method for performing a diagnosis on an individual with a thyroid nodule, comprising the steps of:

- i) extracting RNA from a sample collected from the thyroid of an individual,
- 5 ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21 or from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and
- 10 15 20 iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma.

The invention in a further aspect relates to a method for diagnosing if an individual has, or is at risk of developing, follicular thyroid carcinoma or fetal adenoma, comprising the steps of:

- 25 i) extracting RNA from a sample collected from the thyroid of an individual,
- ii) analysing the mRNA expression profile of the sample, comprising six or more mRNA selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11,
- 30

5 KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,

wherein a predetermined mRNA expression profile of the said mRNAs is indicative of the individual having, or being at risk of developing, follicular thyroid carcinoma or fetal adenoma; said predetermined mRNA expression profile being associated with a prediction according to the mRNA classifier disclosed herein.

10

The invention in a further aspect relates to a method for determining the need for thyroidectomy in an individual presenting with a thyroid nodule, comprising the steps of:

- i) extracting RNA from a sample collected from the thyroid of an individual,
- 15 ii) analysing the mRNA expression profile of the sample, comprising six or more selected from the group disclosed in tables 19, 20 and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, 20 BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSP1, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, 25 SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
- iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma, and
- 30 iv) performing thyroidectomy on the individual only if the nodule is diagnosed as follicular thyroid carcinoma or fetal adenoma,

as determined according to the mRNA classifier disclosed herein.

The invention in a further aspect relates to a method for avoiding thyroidectomy in an individual presenting with a thyroid nodule, comprising the steps of:

- i) extracting RNA from a sample collected from the thyroid of an individual,
- 5 ii) analysing the mRNA expression profile of the sample, comprising six or more mRNA selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, 10 BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, 15 SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
- iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma, and
- 20 iv) performing thyroidectomy on the individual only if the nodule is diagnosed as follicular thyroid carcinoma or fetal adenoma,

as determined according to the mRNA classifier disclosed herein.

25 The invention in a further aspect relates to a method for partitioning a group of individuals presenting with thyroid nodules, comprising the steps of:

- i) extracting RNA from a sample collected from the thyroid of an individual,
- ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, 30 DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B,

5 BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,

10 iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma, and

iv) performing thyroidectomy on the group of individuals only on thyroid nodules diagnosed as follicular thyroid carcinoma or fetal adenoma,

as determined according to the mRNA classifier disclosed herein.

15 The invention in a further aspect relates to a method for performing thyroidectomy in an individual presenting with a thyroid nodule, comprising the steps of:

i) extracting RNA from a sample collected from the thyroid of an individual,

20 ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, 25 RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, 30 JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,

iii) determining if said individual has a malignant or pre-malignant condition selected from follicular thyroid carcinoma or fetal adenoma.

- iv) performing thyroidectomy on the individual if the nodule is diagnosed as follicular thyroid carcinoma or fetal adenoma,

as determined according to the mRNA classifier disclosed herein.

- 5 It follows, that any of the above-mentioned methods may comprise the step of obtaining prediction probabilities of between 0-1.

In a further embodiment, any of the above-mentioned methods may be used in combination with at least one additional diagnostic method.

10

Said at least one additional diagnostic method may in one embodiment be selected from the group consisting of Scintillation counting, Blood sample analysis, Ultrasound imaging, Cytology, Histology and Assessment of risk factors. These are described herein above.

15

In a preferred embodiment, said at least one additional diagnostic method improves the sensitivity and/or specificity of the combined diagnostic outcome.

20

The invention in a further aspect relates to a method for expression profiling of a sample, comprising measuring at six or more mRNAs selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and correlating said expression profile to a clinical condition.

30

In one embodiment, said clinical condition is follicular thyroid carcinoma, follicular thyroid adenoma or fetal adenoma.

35

In another aspect, the present invention relates to a method for determining the prognosis of an individual with a thyroid nodule, comprising the steps of

- i) extracting RNA from a sample collected from the thyroid of an individual,
- 5 ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, 10 BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, 15 SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
- iii) determining if said individual has a malignant or pre-malignant condition selected from follicular thyroid carcinoma or fetal adenoma.

as determined according to the mRNA classifier disclosed herein.

20

Sample type

The sample according to the present invention is extracted from an individual and used for mRNA profiling for the subsequent diagnosis of a condition. In one embodiment, the sample comprises cells and/or tissue.

25

The sample may be collected from an individual or a cell culture, preferably an individual. The individual may be any animal, such as a mammal, including human beings. In a preferred embodiment, the individual is a human being.

30

In a particular embodiment, the sample is taken from the thyroid gland of a human being, such as a thyroid gland comprising thyroid neoplasia and/or a thyroid nodule. In particular, the sample comprises tumour cells, if present. In a preferred embodiment, the sample is obtained from a thyroid nodule of an individual.

Sample collection

5 In one embodiment, the sample is collected from the thyroid of an individual by any available means, such as fine-needle aspiration (FNA) using a needle with a maximum diameter of 1 mm; core needle aspiration using a needle with a maximum diameter of above 1 mm (also called coarse needle aspiration or biopsy, large needle aspiration or large core aspiration); cutting biopsy; open biopsy; a surgical sample; or any other means known to the person skilled in the art. In another embodiment, the sample is collected from an in vitro cell culture.

10 In a preferred embodiment, the sample is a fine-needle aspirate from an individual. The fine-needle aspiration may be performed using a needle with a diameter of between 0.2 to 1.0 mm, such as 0.2 to 0.3 mm, for example 0.3 to 0.4 mm, such as 0.4 to 0.5 mm, for example 0.5 to 0.6 mm, such as 0.6 to 0.7 mm, for example 0.7 to 0.8 mm, such as 0.8 to 0.9 mm, for example 0.9 to 1.0 mm in diameter.

15 The sample may in one preferred embodiment be extracted by the method disclosed in co-pending patent application filed on the same date as the present patent application, with agents reference number P 2038 DK00 titled 'Improved RNA purification method'.

20 The diameter of the needle is indicated by the needle gauge. Various needle lengths are available for any given gauge. Needles in common medical use range from 7 gauge (the largest) to 33 (the smallest) on the Stubbs scale. Although reusable needles remain useful for some scientific applications, disposable needles are far more common in medicine. Disposable needles are embedded in a plastic or aluminium hub
25 that attaches to the syringe barrel by means of a press-fit (Luer) or twist-on (Luer-lock) fitting.

The fine-needle aspiration is in a preferred embodiment performed using a needle gauge of between 20 to 33, such as needle gauge 20, for example needle gauge 21,
30 such as needle gauge 22, for example needle gauge 23, such as needle gauge 24, for example needle gauge 25, such as needle gauge 26, for example needle gauge 27, such as needle gauge 28, for example needle gauge 29, such as needle gauge 30, for example needle gauge 31, such as needle gauge 32, for example needle gauge 33. In a particular embodiment, the gauge of the needle is 23.

35

The fine-needle aspiration may in one embodiment be assisted, such as ultra-sound (US) guided fine-needle aspiration, x-ray guided fine-needle aspiration, endoscopic ultra-sound (EUS) guided fine-needle aspiration, Endobronchial ultrasound-guided fine-needle aspiration (EBUS), ultrasonographically guided fine-needle aspiration, 5 stereotactically guided fine-needle aspiration, computed tomography (CT)-guided percutaneous fine-needle aspiration and palpation guided fine-needle aspiration.

The skin above the area to be biopsied may in one embodiment be swiped with an antiseptic solution and/or may be draped with sterile surgical towels. The skin, 10 underlying fat, and muscle may in one embodiment be numbed with a local anesthetic.

After the needle is placed into the mass, cells may be withdrawn by aspiration with a syringe.

15 The sample extracted from an individual by any means as disclosed above may be transferred to a tube or container prior to analysis. The container may be empty, or may comprise a collection media. Collection media are disclosed herein below.

The sample extracted from an individual by any means as disclosed above may be 20 analysed essentially immediately, or it may be stored prior to analysis for a variable period of time and at various temperature ranges.

In one embodiment, the sample is stored at a temperature of between -200°C to 37°C, such as between -200 to -100°C, for example -100 to -50°C, such as -50 to -25°C, for 25 example -25 to -10°C, such as -10 to 0°C, for example 0 to 10°C, such as 10 to 20°C, for example 20 to 30°C, such as 30 to 37°C prior to analysis.

In another embodiment, the sample is stored for between 15 minutes and 100 years prior to analysis, such as between 15 minutes and 1 hour, for example 1 to 2 hours, 30 such as 2 to 5 hours, for example 5 to 10 hours, such as 10 to 24 hours, for example 24 hours to 48 hours, such as 48 to 72 hours, for example 72 to 96 hours, such as 4 to 7 days, such as 1 week to 2 weeks, such as 2 to 4 weeks, such as 4 weeks to 1 month, such as 1 month to 2 months, for example 2 to 3 months, such as 3 to 4 months, for example 4 to 5 months, such as 5 to 6 months, for example 6 to 7 months, such as 7 to 8 35 months, for example 8 to 9 months, such as 9 to 10 months, for example 10 to 11

moths, such as 11 to 12 months, for example 1 year to 2 years, such as 2 to 3 years, for example 3 to 4 years, such as 4 to 5 years, for example 5 to 6 years, such as 6 to 7 years, for example 7 to 8 years, such as 8 to 9 years, for example 9 to 10 years, such as 10 to 20 years, for example 20 to 30 years, such as 30 to 40 years, for example 40 to 50 years, such as 50 to 75 years, for example 75 to 100 years prior to analysis.

Collection media for sample

A collection media according to the present invention is any solution suitable for collecting a sample for immediate or later analysis and/or retrieval of RNA from said sample.

In one embodiment, the collection media is an RNA preservation solution or reagent suitable for containing samples without the immediate need for cooling or freezing the sample, while maintaining RNA integrity prior to extraction of RNA from the sample. An RNA preservation solution or reagent may also be known as RNA stabilization solution or reagent or RNA recovery media, and may be used interchangeably herein. The RNA preservation solution may penetrate the harvested cells of the collected sample and retards RNA degradation to a rate dependent on the storage temperature.

The RNA preservation solution may be any commercially available solutions or it may be a solution prepared according to available protocols.

The commercially available RNA preservation solutions may for example be selected from RNeasy® (Qiagen), RNeasy Lysis Buffer (Qiagen), RNeasy Protect Tissue Reagent (Qiagen) and RNeasy Protect Cell Reagent (Qiagen). Protocols for preparing a RNA stabilizing solution may be retrieved from the internet (e.g. L.A. Clarke and M.D. Amaral: 'Protocol for RNase-retarding solution for cell samples', provided through The European Working Group on CFTR Expression), or may be produced and/or optimized according to techniques known to the skilled person.

In another embodiment, the collection media will penetrate and lyse the cells of the sample immediately, including reagents and methods for isolating RNA from a sample that may or may not include the use of a spin column.

Said reagents and methods for isolating RNA is described herein below in the section 'analysis of sample'.

Other collection media according to the present invention comprises any media such as water, sterile water, denatured water, saline solutions, buffers, PBS, TBS, Allprotect Tissue Reagent (Qiagen), cell culture media such as RPMI-1640, DMEM (Dulbecco's Modified Eagle Medium), MEM (Minimal Essential Medium), IMDM (Iscove's Modified Dulbecco's Medium), BGjB (Fitton-Jackson modification), BME (Basal Medium Eagle), Brinster's BMOC-3 Medium, CMRL Medium, CO₂-Independent Medium, F-10 and F-12 Nutrient Mixture, GMEM (Glasgow Minimum Essential Medium), IMEM (Improved Minimum Essential Medium), Leibovitz's L-15 Medium, McCoy's 5A Medium, MCDB 131 Medium, Medium 199, Opti-MEM, Waymouth's MB 752/1, Williams' Media E, Tyrode's solution, Belyakov's solution, Hanks' solution and other cell culture media known to the skilled person, tissue preservation media such as HypoThermosol®, CryoStor™ and Steinhardt's medium and other tissue preservation media known to the skilled person.

Sample analysis

After the sample is collected, it is subjected to analysis. In one embodiment, the sample is initially used for isolating or extracting RNA according to any conventional methods known in the art; followed by an analysis of the mRNA expression in said sample.

Extraction of RNA

The RNA isolated from the sample may be total RNA, or mRNA.

Conventional methods and reagents for isolating RNA from a sample comprise Trizol (Invitrogen), Guanidinium thiocyanate-phenol-chloroform extraction, PureLink Micro-to-Midi Total RNA Purification System (Invitrogen), RNeasy kit (Qiagen), Oligotex kit (Qiagen), phenol extraction, phenol-chloroform extraction, TCA/acetone precipitation, ethanol precipitation, Column purification, Silica gel membrane purification, PureYield™ RNA Midiprep (Promega), PolyATtract System 1000 (Promega), Maxwell® 16 System (Promega), SV Total RNA Isolation (Promega), geneMAG-RNA / DNA kit (Chemicell), TRI Reagent® (Ambion), RNAqueous Kit (Ambion), ToTALLY RNA™ Kit (Ambion), Poly(A)Purist™ Kit (Ambion) and any other methods, commercially available or not, known to the skilled person.

The RNA may be further amplified, cleaned-up, concentrated, DNase treated, quantified or otherwise analysed or examined such as by agarose gel electrophoresis or Bioanalyser analysis (Agilent) or subjected to any other post-extraction method known to the skilled person.

Methods for extracting and analysing an RNA sample are disclosed in Molecular Cloning, A Laboratory Manual (Sambrook and Russell (ed.), 3rd edition (2001), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Microarray analysis

The isolated RNA is in one preferred embodiment analysed by microarray analysis.

A microarray is a multiplex technology that consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides or antisense mRNA probes, called features, each containing picomoles of a specific oligonucleotide sequence. This can be a short section of a gene or other DNA or RNA element that are used as probes to hybridize a DNA or RNA sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by fluorescence-based detection of fluorophore-labeled targets to determine relative abundance of nucleic acid sequences in the target.

In standard microarrays, the probes are attached to a solid surface by a covalent bond to a chemical matrix (via epoxy-silane, amino-silane, lysine, polyacrylamide or others). The solid surface can be glass or a silicon chip, in which case they are commonly known as gene chip. DNA arrays are so named because they either measure DNA or use DNA as part of its detection system. The DNA probe may however be a modified DNA structure such as LNA (locked nucleic acid).

In one embodiment, the microarray analysis as used herein is used to detect mRNA, expression profiling.

The microarray for detection of mRNA may be an microarray platform, wherein the probes of the microarray may be comprised of antisense mRNAs or DNA oligonucleotides. In the first case, the target is a labelled sense mRNA sequence, and in the latter case the mRNA has been reverse transcribed into cDNA and labelled.

The microarray for detection of mRNA may be any commercially available array platform.

- 5 Microarray analysis may comprise all or a subset of the steps of RNA isolation, RNA amplification, reverse transcription, target labelling, hybridisation onto a microarray chip, image analysis and normalisation, and subsequent data analysis; each of these steps may be performed according to a manufacturers protocol such as Invitrogen, or as described herein below in Example 1.

10

It follows, that any of the methods as disclosed herein above e.g. for diagnosing of an individual with a thyroid nodule may further comprise one or more of the steps of:

- i) isolating mRNA from a sample,
- ii) labelling of said mRNA,
- 15 iii) hybridising said labelled mRNA to a microarray comprising mRNA-specific probes to provide a hybridisation profile for the sample,
- iv) performing data analysis to obtain a measure of the mRNA expression profile of said sample.

- 20 In another embodiment, the microarray for detection of mRNA is custom made.

A probe or hybridization probe is a fragment of DNA or RNA of variable length, which is used to detect in DNA or RNA samples the presence of nucleotide sequences (the target) that are complementary to the sequence in the probe. One example is a sense
25 mRNA sequence in a sample (target) and an antisense mRNA probe. The probe thereby hybridizes to single-stranded nucleic acid (DNA or RNA) whose base sequence allows probe-target base pairing due to complementarity between the probe and target.

- 30 To detect hybridization of the probe to its target sequence, the probe or the sample is tagged (or labeled) with a molecular marker. Detection of sequences with moderate or high similarity depends on how stringent the hybridization conditions were applied — high stringency, such as high hybridization temperature and low salt in hybridization buffers, permits only hybridization between nucleic acid sequences that are highly
35 similar, whereas low stringency, such as lower temperature and high salt, allows

hybridization when the sequences are less similar. Hybridization probes used in microarrays refer to nucleotide sequences covalently attached to an inert surface, such as coated glass slides, and to which a mobile target is hybridized. Depending on the method the probe may be synthesised via phosphoramidite technology or generated by PCR amplification or cloning (older methods). To design probe sequences, a probe design algorithm may be used to ensure maximum specificity (discerning closely related targets), sensitivity (maximum hybridisation intensities) and normalised melting temperatures for uniform hybridisation.

10 *RT-QPCR*

In another embodiment, the isolated RNA is analysed by quantitative ('real-time') PCR (QPCR).

Real-time polymerase chain reaction, also called quantitative polymerase chain reaction (Q-PCR/qPCR/RT-QPCR) or kinetic polymerase chain reaction, is a technique based on the polymerase chain reaction, which is used to amplify and simultaneously quantify a targeted DNA molecule. It enables both detection and quantification (as absolute number of copies or relative amount when normalized to DNA input or additional normalizing genes) of a specific sequence in a DNA sample.

20

The procedure follows the general principle of polymerase chain reaction; its key feature is that the amplified DNA is quantified as it accumulates in the reaction in *real time* after each amplification cycle. Two common methods of quantification are the use of fluorescent dyes that intercalate with double-stranded DNA, and modified DNA oligonucleotide probes that fluoresce when hybridized with a complementary DNA.

25

Frequently, real-time polymerase chain reaction is combined with reverse transcription polymerase chain reaction to quantify low abundance messenger RNA (mRNA), enabling a researcher to quantify relative gene expression at a particular time, or in a particular cell or tissue type.

30

In a real time PCR assay a positive reaction is detected by accumulation of a fluorescent signal. The Ct (cycle threshold) is defined as the number of cycles required for the fluorescent signal to cross the threshold (i.e. exceeds background level). Ct levels are inversely proportional to the amount of target nucleic acid in the sample (i.e.

the lower the Ct level the greater the amount of target nucleic acid in the sample). Most real time assays undergo 40 cycles of amplification.

5 Cts < 29 are strong positive reactions indicative of abundant target nucleic acid in the sample. Cts of 30-37 are positive reactions indicative of moderate amounts of target nucleic acid. Cts of 38-40 are weak reactions indicative of minimal amounts of target nucleic acid which could represent an infection state or environmental contamination.

10 The QPCR may be performed using chemicals and/or machines from a commercially available platform.

The QPCR may be performed using chemicals and/or machines from a commercially available platform.

15 The QPCR may be performed using QPCR machines from any commercially available platform; such as Prism, geneAmp or StepOne Real Time PCR systems (Applied Biosystems), LightCycler (Roche), RapidCycler (Idaho Technology), MasterCycler (Eppendorf), iCycler iQ system, Chromo 4 system, CFX, MiniOpticon and Opticon systems (Bio-Rad), SmartCycler system (Cepheid), RotorGene system (Corbett
20 Lifescience), MX3000 and MX3005 systems (Stratagene), DNA Engine Opticon system (Qiagen), Quanta qPCR systems (Techne), InSyte and Syncrom cyclor system (BioGene), DT-322 (DNA Technology), Exicycler Notebook Thermal cyclor, TL998 System (Ilanlong), Line-Gene-K systems (Bioer Technology), or any other commercially available platform.

25 The QPCR may be performed using chemicals from any commercially available platform, such as NCode EXPRESS qPCR or EXPRESS qPCR (Invitrogen), Taqman or SYBR green qPCR systems (Applied Biosystems), Real-Time PCR reagents (Eurogentec), iTaq mix (Bio-Rad), qPCR mixes and kits (Biosense), and any other
30 chemicals, commercially available or not, known to the skilled person.

The QPCR reagents and detection system may be probe-based, or may be based on chelating a fluorescent chemical into double-stranded oligonucleotides.

The QPCR reaction may be performed in a tube; such as a single tube, a tube strip or a plate, or it may be performed in a microfluidic card in which the relevant probes and/or primers are already integrated.

- 5 A Microfluidic card allows high throughput, parallel analysis of mRNA expression patterns, and allows for a quick and cost-effective investigation of biological pathways. The microfluidic card may be a piece of plastic that is riddled with micro channels and chambers filled with the probes needed to translate a sample into a diagnosis. A sample in fluid form is injected into one end of the card, and capillary action causes the
- 10 fluid sample to be distributed into the microchannels. The microfluidic card is then placed in an appropriate device for processing the card and reading the signal.

It follows, that any of the methods as disclosed herein above e.g. for diagnosing of an individual with a thyroid nodule may further comprise one or more of the steps of:

- 15 i) isolating mRNA from a sample,
ii) performing QPCR analysis on said mRNA,
iii) performing data analysis to obtain a measure of the mRNA expression profile of said sample.

20 *Other analysis methods*

In yet another embodiment, the isolated RNA is analysed by northern blotting.

- A northern blot is a method used to check for the presence of a RNA sequence in a sample. Northern blotting combines denaturing agarose gel or polyacrylamide gel
- 25 electrophoresis for size separation of RNA with methods to transfer the size-separated RNA to a filter membrane for probe hybridization. The hybridization probe may be made from DNA or RNA.

- In yet another embodiment, the isolated RNA is analysed by nuclease protection
- 30 assay.

- Nuclease protection assay is a technique used to identify individual RNA molecules in a heterogeneous RNA sample extracted from cells. The technique can identify one or more RNA molecules of known sequence even at low total concentration. The
- 35 extracted RNA is first mixed with antisense RNA or DNA probes that are

complementary to the sequence or sequences of interest and the complementary strands are hybridized to form double-stranded RNA (or a DNA-RNA hybrid). The mixture is then exposed to ribonucleases that specifically cleave only *single*-stranded RNA but have no activity against double-stranded RNA. When the reaction runs to completion, susceptible RNA regions are degraded to very short oligomers or to individual nucleotides; the surviving RNA fragments are those that were complementary to the added antisense strand and thus contained the sequence of interest.

10 Device

It is also an aspect of the present invention to provide a device for measuring the expression level of six or more mRNAs, wherein said device comprises or consists of probes for mRNAs selected from the group disclosed in tables 19, 20 and 21 or the group consisting of

15 FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15,
20 COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,

25 and wherein said device is used for classifying a sample obtained from a thyroid nodule of an individual.

In another embodiment, the device comprises or consists of probes for mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1,
30 LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91,

C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, and ZNF330.

5 In a further embodiment, the device comprises or consists of probes for mRNAs selected from the group consisting of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, 10 and SLCO2A1.

In yet another embodiment, the device comprises or consists of probes selected for mRNAs from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.

15

It is understood, that probes may be a single probe or a probe set comprising multiple probes for the same mRNA.

20 In one embodiment, the device may be used for distinguishing between thyroid follicular adenoma and thyroid follicular carcinoma, and/or distinguishing thyroid follicular adenoma and thyroid follicular carcinoma and fetal adenoma, and/or distinguishing thyroid follicular carcinoma and thyroid follicular adenoma merged with fetal adenoma.

25 In one embodiment said device comprises between 1 to 5 probes and/or probe sets for one or more mRNAs according to the present invention, such as 5 to 10, for example 10 to 15, such as 15 to 20, for example 20 to 25, such as 25 to 30, for example 30 to 35, such as 35 to 40, for example 40 to 45, such as 45 to 50, for example 50 to 55, such as 55 to 60, for example 60 to 65, such as 65 to 70, for example 70 to 75, such as 30 75 to 80, for example 80 to 85, such as 85 to 90, for example 90 to 95, such as 95 to 100, for example 100 to 110, such as 110 to 120 probes and/or probe sets for one or mroe mRNA according to the present invention.

35 In a particular embodiment, said device comprises 66 probes and/or probe sets. In a particular embodiment, said device comprises probes and/or probe sets for 64 mRNAs.

In another embodiment said device comprises less than 120 probes and/or probe sets for a mRNA according to the present invention, such as less than 110 probes, for example less than 100 probes, such as less than 90 probes, for example less than 80 probes, such as less than 70 probes, for example less than 60 probes, such as less than 50 probes, for example less than 40 probes, such as less than 30 probes, for example less than 20 probes, such as less than 10 probes and/or probe sets for a mRNA according to the present invention.

10 In one embodiment, the device may be a microarray chip comprising six or more probes for a mRNA selected from the group disclosed in tables 19, 20 and 21 or FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, 15 HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, 20 TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.

In another embodiment, the device may be a QPCR Micro Fluidic Card comprising six or more probes for a mRNA selected from the group disclosed in tables 19, 20 and 21 25 or FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, 30 COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.

35

In yet another embodiment, the device may comprise QPCR tubes, QPCR tubes in a strip or a QPCR plate comprising six or more probes for a mRNA selected from the group disclosed in tables 19, 20 and 21 or FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.

The probes may be comprised on a solid support, on at least one bead, or in a liquid reagent comprised in a tube.

Computer program product

It is a further aspect of the invention to provide a computer program product having a computer readable medium, said computer program product comprising means for carrying out any of the herein listed mRNA classifiers, models and methods.

It is a further aspect of the invention to provide a system comprising means for carrying out any of the herein listed methods.

It is an aspect of the present invention to provide a system for determining the presence of a malignant and/or pre-malignant condition in a sample obtained from a thyroid nodule of an individual, said system comprising means for analysing the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the group disclosed in tables 19, 20 and 21 or the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD,

CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,

5 wherein said expression level of said mRNAs is associated with thyroid follicular carcinoma, thyroid follicular adenoma merged with fetal adenoma, or thyroid follicular carcinoma or fetal adenoma, said association being predicted according to the mRNA classifier disclosed herein.

10 It is a further aspect of the present invention to provide a system for determining the presence of a benign condition in a sample obtained from a thyroid nodule of an individual, said system comprising means for analysing the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the group disclosed in tables 19, 20 and 21 or the group consisting of FOSB, LOC286002, CA4,
 15 EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB,
 20 C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
 wherein said expression level of said mRNAs is associated with thyroid follicular
 25 adenoma, said association being predicted according to the mRNA classifier disclosed herein.

In another aspect, the present invention provides a system for performing a diagnosis on an individual with a thyroid nodule, comprising:

30 i) means for analysing the mRNA expression profile of the thyroid nodule, comprising six or more mRNAs selected from the group disclosed in tables 19, 20 and 21 or the group of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C,

CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and

ii) means for determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma.

In another aspect, the present invention provides a computer program product having a computer readable medium, said computer program product providing a system for predicting the diagnosis of an individual with a thyroid nodule, said computer program product comprising means for carrying out any of the steps of any of the methods as disclosed herein.

In another aspect, the present invention provides a system as disclosed herein wherein the data is stored, such as stored in at least one database.

Kit-of-parts

It is also an aspect to provide a kit-of-parts comprising the device according to the present invention, and at least one additional component.

In one embodiment, said additional component is means for extracting RNA such as mRNA from a sample; reagents for performing microarray analysis and/or reagents for performing QPCR analysis.

In another embodiment, said kit may comprise instructions for use of the device and/or the additional components.

In a further embodiment, said kit comprises a computer program product having a computer readable medium as detailed herein elsewhere.

Detailed description of the drawings

Figure 1. PCA on 92 different thyroid samples based on the entire transcriptome (all genes on the array). Each dot represents a sample that is colored according to the histological grouping. Follicular adenoma (FA) – light blue, follicular carcinoma (FC) - orange, fetal adenoma (FEA) - green, anaplastic carcinoma (AC) - red, papillary carcinoma (PC) - dark blue, nodular goiter (NG) - pink and normal thyroid (NT) - white. The PCA plot of the first three principal components captures 27% of the variance across the samples. The percentage of the total variance described by each of the three principal components is shown in parenthesis at each axis. The AC group and the NG are both well separated from the remaining samples. FA, FC, and FEA are difficult to distinguish, although FEA are mainly localized in the lower part of the cluster. The FC nodules are very heterogeneous, showing similarity with PC, FA and FEA. The PCA analysis and visualization was performed using the Qlucore Gene Expression Explorer 1.1 (www.qcore.com).

Figure 2. Functional categories of genes that are differentially expressed between follicular adenoma and carcinoma. A total of 117 probe sets were selected as being differentially expressed with Benjamini Hochberg's corrected p values below 0.05, absolute fold changes above 1.5 and an absolute change in expression levels of more than 100. The probe sets were pruned to remove redundant probe set, which resulted in a list of 101 unique genes regulated genes. The putative role of every differentially expressed factor was derived from the comprehensive cDNA-supported gene and transcripts annotation Ace View (Genome Biology 2006) (www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.htm) and substantiated by a search in PubMed and Google. The genes were grouped according to the functional categories; DNA-RNA binding, cell cycle and apoptosis, metabolism, signaling, ECM and cytoskeleton, protein secretion and unknown function. The percentage of genes in each category is shown.

Figure 3. List of gene products related to cell cycle control, specifically s-phase and mitosis, and apoptosis, which were found to be highly enriched within the genes that are regulated between follicular adenoma and –carcinoma. The genes in the two categories are shown along with fold change (FC) and P-value. Positive and negative fold changes represents genes that are up and down regulated in follicular carcinoma, respectively.

Figure 4A. An Immunostaining of follicular adenoma (FA) and follicular carcinoma (FC). The upper panels are FA, and the lower panels are FC. Sections of formalin fixed paraffin embedded tissues were stained with Hematoxylin and Eosin for pathological study (the left panels). Immunostainings with antibodies against Ki67, TOP2A, NR4A1, and NR4A3 were shown, respectively. The scale bar is 20 um and the original magnification was x400.

Figure 4B. Hierarchical cluster visualization of genes involved in cell cycle, mitosis and apoptosis. The genes were clustered using rank correlation and average linkage clustering. Each row represents a gene and each column a sample. Follicular adenoma samples are labeled in grey and follicular carcinoma in brown, where A denotes adenoma and C denotes carcinoma. The color bar indicate the degree of up or down regulation, with dark blue representing a down regulation of two standard deviations from the mean and red represents up regulation of two standard deviations or more from the mean. The heat map illustrates a loss of apoptotic and growth arrest factors during transformation to malignancy, i.e. up-regulated in FA samples and down in FC. Up-regulation of cell-cycle associated transcripts in FC that was graduate compared to the loss of apoptotic factors, which was observed in all samples.

Figure 5. Bar graph of average expression level of deregulated apoptotic and mitotic factors in benign and malignant samples. Blue graphs show apoptotic factors that are up-regulated in follicular and anaplastic carcinoma, and red graphs show mitotic factors, which are down-regulated in cancer. NT – normal thyroid, NG – nodular goiter, FA – follicular adenoma, FEA – fetal adenoma, FC – follicular carcinoma, AC – anaplastic carcinoma, PC – papillar carcinoma.

Figure 6A. Principal component analysis (PCA) of 76 probe sets defined in the FA versus FC classifier. The PCA analysis and visualization was performed by using the Qlucore Gene Expression Explorer 1.1 (www.qlucore.com). Each dot represents a sample, FA - orange, FC - blue. The three dimensional PCA plot captures 72% of the variance described by the gene signature. The percentage of the total variance described by each of the three principal components is shown in parenthesis at each axis. The FA and FC groups are well separated, except for two samples (FC8 and FA18), which are located in the borderline area between the FA and FC clusters.

These samples are misclassified in the classification analysis. The PCA analysis and visualization was performed by using the Qlucore Gene Expression Explorer 1.1 (www.qlucore.com).

5 Figure 6B. Plot of the predictive Probability output for classifier 1. The first 40 samples constitute the training set of 22 FA samples and 18 FC samples. Samples 41 through 52 are included as test set, which consists of 12 FEA samples. If a sample has a predictive value above 0.5 ($p(\text{FA}) > 0.5$), it is classified as follicular adenoma, otherwise as a follicular carcinoma. Each dot represents a sample and the color show the true class. FA – orange, FC – blue and FEA – green. The samples FC8 and FA18 from the training set are misclassified during cross validation. Only one FEA sample, FEA9, is classified as a follicular carcinoma.

15 Figure 6C. Receiver operating characteristic (ROC) curve for the binary classifier built to distinguish between follicular adenoma and –carcinoma nodules. The curve shows the true positive rate versus false positive rate i.e. the tradeoff between sensitivity and specificity. The area under the curve (AUC), which captures the ability of the classifier to correctly group those patients with follicular adenoma and those with follicular carcinoma, is equal to 0.96. A perfect classifier will have an AUC of 1.0, whereas an AUC value of 0.5 indicates that the classification is random.

25 Figure 7A. Principal component analysis (PCA) of the 116 signatures defined in classifier 2 built to separate the three nodules, FA, FEA and FC. Each dot represents a sample, follicular adenoma (FA) – light blue, follicular carcinoma (FC) – orange and fetal adenoma (FEA) – green. The three dimensional PCA plot captures 67% of the variance described by the gene signature. The percentage of the total variance described by each of the three principal components is shown in parenthesis at each axis. All groups can be easily distinguished from each other, except for a few samples that are the FA and FC groups are well separated, except for a few samples, which are located in the borderline area between the FA and FC clusters. The FEA samples overlap with the FA samples. The PCA analysis and visualization was performed using the Qlucore Gene Expression Explorer 1.1 (www.qlucore.com).

35 Figure 7B. Ternary plot of the predictive probability output obtained for classifier 2. Each dot is a sample, where the color code represents the true class, follicular

adenoma (FA) – light blue, follicular carcinoma (FC) – orange and fetal adenoma (FEA) – green. Each vertex of the triangle represents a sub class. Samples plotted close to a vertex have a high probability of belonging to this particular class, and samples plotted in the center are fully uncertain. The dotted lines indicate the predictive probability with respect to each of the three groups. Using the SVM model as a probabilistic classifier, eight samples are misclassified: four FEA, two FA and two FC samples. The misclassified samples are labeled.

Figure 8. Principal component analysis (PCA) of the 76 probe set signature defined in the FA versus FC classifier. Each dot represents a sample, FA – orange, FC – blue and FEA – green. The three dimensional PCA plot captures 72% of the variance described by the gene signature. The FEA samples were not active in the PCA analysis, rather they are visualized to show how they are grouped using this classifier. The percentage of the total variance described by each of the three principal components is shown in parenthesis at each axis. The FA and FC groups are well separated, except for four samples, two FEA and one follicular adenomas and carcinomas, respectively.

Fig 9A. Principal component analysis (PCA) of the 65 probe set signature defined in classifier 3 (FA and FEA merged versus FC). Each dot represents a sample, adenoma – orange, carcinoma – blue. The three dimensional PCA plot captures 74% of the variance described by these genes. The percentage of the total variance described by each of the three principal components is shown in parenthesis at each axis. The groups are well separated, although five samples are miss-classified during LOOCV training; three adenoma (FA-FEA) samples and two FC samples. The follicular carcinoma cluster is more heterogeneous than the adenoma samples showing the progressive nature of the cancer samples.

Fig 9B. Plot of the predictiveProbability output obtained for classifier 3. The classifier was trained to distinguish between the merged group of fetal adenomas and follicular adenomas versus follicular carcinomas. If a sample has a predictive value above 0.5 ($p(\text{fa}) > 0.5$), it is classified as follicular adenoma, otherwise a follicular carcinoma. Each dot represents a sample where the color code shows the true class. Adenomas – orange and follicular carcinoma – blue. Five samples were miss-classified, three adenoma (FA-FEA) samples and two FC samples.

Examples

Example 1: mRNA classifiers for thyroid follicular neoplasia

To explore if it was possible to identify molecular pathways implicated in follicular neoplasia and further improve diagnosis, we performed a global expression profiling of follicular nodules and applied supervised learning by support vector machines to generate a diagnostic signatures based on the major cancer specific changes. We report that thyroid follicular carcinomas are characterized by transcripts encoding factors involved in DNA replication and mitosis and loss of growth-arrest, and proapoptotic factors such as NR4A1 and NR4A3, FOSB and JUN, which previously have been causally associated to stem cell proliferation and defective extrinsic apoptotic signaling (Mullican *et al.* 2007). Based on the analysis of differentially expressed transcripts, we generated a molecular classifier that could identify carcinomas with high accuracy. Validation employing public-domain and cross-platform data, demonstrated that the signature was robust and worked equally well on follicular nodules originating from different geographical locations and platforms.

Materials and methods

Collection of tumor samples

Sixty-nine (69) tumor samples were collected from patients who underwent thyroidectomy at Copenhagen University Hospital, Rigshospitalet and Odense University Hospital from 1989 to 2008. The sampling at Copenhagen university Hospital was part of an ongoing quality assurance programme, and all patients had been informed about and agreed to do the sampling. All handling and usage of the samples from Odense University Hospital was approved by the Ethics Committee of the County of Funen. After surgical excision, the tumor samples were snap frozen in liquid nitrogen and stored at -80 °C. The tumors included 22 benign follicular adenomas (7 from Copenhagen and 15 from Odense), 18 follicular carcinomas (3 from Copenhagen and 15 from Odense), 12 samples of microfollicular adenomas (all obtained from Odense), 4 anaplastic carcinomas (from Copenhagen), 2 papillary carcinomas (from Copenhagen), and 9 nodular goiters (from Copenhagen). 23 samples were obtained from the expression profile repository, Array Express (<http://www.ebi.ac.uk/arrayexpress/>), these counted 14 papillary carcinomas and 9 normal thyroid.

Microarray analysis

Total RNA was isolated by TRIzol™ reagent (Invitrogen) and purified over RNeasy Columns (Qiagen). The quantity and integrity of the extracted RNA were determined by Nanodrop (Nanodrop technologies) and the Bioanalyzer LabChips (Agilent Technologies), respectively. Samples were labeled according to the manufactures guidelines. In short, 2 µg of total RNA was transcribed into cDNA using an oligo-dT primer containing a T7 RNA polymerase promoter. cDNA was used as a template in the *in-vitro*-transcription reaction driven by the T7 promoter under which, biotin labeled oligo-nucleotides were incorporated into the synthesized cRNA. The labeled cRNAs were hybridized to the HG-U133plus2 GeneChip array (Affymetrix, Santa Clara, CA, USA), which query close to 48,000 well substantiated genes by approximately 56,000 probe sets. The arrays were washed and stained with phycoerythrin conjugated streptavidin (SAPE) using the Affymetrix Fluidics Station® 450, and scanned in the Affymetrix GeneArray® 3000 7G scanner to generate fluorescent images, as described in the Affymetrix GeneChip® protocol.

Microarray Data analysis

Cel files were imported into the statistical software package R v. 2.7.2 using BioConductor v. 2.8 (Gentleman *et al.* 2004) and gcRMA modeled using quantiles normalization and “lowess” summarization (Bolstad *et al.* 2003). The modeled log-intensity of 56,400 probe sets was used for high-level analysis of selecting differentially expressed genes and formulating the classifier. All model construction and optimization were written in R (v. 2.7.2). Various functions from the BioConductor packages, Biobase, affy, multtest, MASS, class, e1071, mda, grid and RocR were applied in the code (Gentleman *et al.* 2004). The microarray data was submitted to the gene expression repository at Array Express (<http://www.ebi.ac.uk/arrayexpress/>) with accession number E-MEXP-2442.

Differential expression analysis

Genes were defined as being differentially expressed in a class comparison analysis if they were selected in the uni-variate two-sample t-test or F-test with equal variance as described below. Statistical hypothesis testing was performed using the multtest package in Bioconductor v. 2.7.2. Firstly, equal variance two-sample t-statistic or multi-sample F-statistic for tests of equality of population means was performed on each gene. Control of Type I error rate was performed by computing adjusted p-values for

simple multiple testing procedures from a vector of raw (unadjusted) p-values. The procedures include the Bonferroni, Holm, Hochberg, and Sidak procedures for strong control of the family-wise Type I error rate (FWER), and the Benjamini & Hochberg and Benjamini & Yekutieli procedures for (strong) control of the false discovery rate (FDR).

5 The FWER methods provide a very conservative control of error rates, and hence the resulting number of rejections (discoveries of differentially expressed genes) is in practice close to zero. In comparison, the FDR methods give more power in the analysis and since we wish to make as many discoveries as possible to enhance the chance of defining the molecular change in the samples, and a small proportion of errors will not change the overall result, we chose to apply the Benjamini & Hochberg FDR analysis (Benjamini Y. *et al.* 1995). A probe set is defined as being differentially expressed if the adjusted p-value is below 0.05 applying Benjamini & Hochberg controlling procedure (Benjamini Y. *et al.* 1995), and have a fold change larger than 1.5 and a difference of means larger than 100 (real unlogged values) between (mutual) 10 classes of samples (FC versus FA, FEA versus FA and FC versus FEA). The differentially expressed genes were grouped according to their functional categories in cell cycle, cytoskeleton and ECM, DNA binding and transcription, metabolism, RNA processing and translation, secretion and signaling.

20 *Formulation of classifier*

The diagnostic classifiers were developed in R v. 2.7. For all classification problems, the training of the classifiers inside the leave-one-out (LOO) loop consists of two steps: a univariate probe ranking and selection step followed by fitting a support vector machine (SVM) on the sample division using the selected probes as covariates. All 25 models were optimized by a grid search of p-value cut-offs and the cut-off resulting in a gene signature of optimal performance was used in the final model. The gene signature in classifier 1 and 3 was selected with Students t-test with p-values below $1e-4$ and $1e-6$, and in classifier 2 with an F test with p-value below $5e-6$. Model fitting was done by training an SVM with a Gaussian kernel (Vapnik 1998). The parameters of the classifier (cost and gamma) were selected by grid search using different combinations of values 30 and cross validation within leave-one-out loops to ensure that the estimation of the classifier parameters was unbiased. The grid search optimization showed that a spectrum of values of cost and gamma provided similar performance, and the median values were used in the algorithm. For each cross-validation loop, the percentage of genes selected is reported, and we applied this measure to enhance the robustness of 35

the model in classifier 1 by using only the probe sets that have 100 percent cross validation support in the final classifier.

5 The trained SVM model was turned into a probabilistic classifier giving an estimate of the probability of the predicted class label, i.e. quantify the prediction uncertainty, or the predictive probability of a sample being one or the other type using logit estimates (Platt JC 1999). The predictive probability is graphed as the function $p(\text{FA})$ by plotting the predictive probability on the y-axis and samples on the x-axis (classifier 1 and 3). In the three-class problem (classifier 2) a ternary plot was produced combining the
10 probabilities for a sample being either one of the three classes, FA, FC or FEA.

Estimation of misclassification rate

The misclassification rate for each classifier was evaluated using leave-one-out cross-validation (LOOCV) during which we applied t-tests (classifier 1 and 3) or the F-test
15 (classifier 2) for feature selection of probe sets to include in each model. The correct classification rate was calculated as the percentage of correctly classified out of the total number of samples examined. Furthermore, the performance of each classifier during LOOCV is described by the following parameters: sensitivity, which is the probability for a class A sample to be correctly predicted as class A; specificity, which is
20 the probability for a none class A sample to be correctly predicted as non-A; positive redictive value (PPV) is the probability that a sample predicted as class A actually belongs to class A, and negative predictive value (NPV), the probability that a sample predicted as none class A actually does not belong to class A (Simon *et al.* 2007).

25 *Statistical significance of the Error rate*

A permutation test was performed in order to determine if the cross-validated misclassification rate is lower than expected by chance (Tusher *et al.* 2001; Simon *et al.* 2007). In 1000 random permutations of the class label, the entire cross-validation was repeated for classifying the random classes of samples. The proportion of the
30 1000 random permutations that gives a smaller or similar cross-validation misclassification rate as obtained with the real data determines the permutation p-value. The statistical significance of the error rate was determined for the SVM classifier in the two class cases, and using the 3-Nearest Neighbors (3-NN) method for the three-class case due to computational limitations.

35

Comparison of different classification models

The performance of the SVM classifier was compared to other classifiers based on different algorithms, these being diagonal linear discriminant analysis (DLDA), compound covariate predictor (CCP), 1-Nearest Neighbor (1-NN) and 3-Nearest
5 Neighbors (3-NN) using BRB-Array tools (Simon *et al.* 2007). Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for all algorithms.

External validation data sets

10 To validate that classifier 1 can distinguish between FA and FC outside the training setting, we included two external data sets (Weber *et al.* 2005; Hinsch *et al.* 2009), which included expression profiles of 24 samples (12 FA and 12 FC) analyzed with Affymetrix HG-U133A arrays and 12 samples (4 FA and 8 FC) analyzed with ABI Human Genome Survey Microarray version 2 from Applied Biosystem, respectively.
15 The raw data files (cel files) from the Weber *et al.* study were normalized and summarized, expression values were calculated using invariant set normalization and the PM-MM model implemented in the dChip software (Li & Wong 2001) as described (Weber *et al.* 2005). We identified the approximately 22.200 probe sets that were shared between the HG-U133A and the HG-U133_plus2 arrays. The normalized data
20 from the Hinsch *et al.* study were downloaded from the expression profile warehouse at the NCBI the Gene expression Omnibus (Barrett *et al.* 2009) GEO accession number "GEO15045" and the unique Ids were coupled with gene name and gene symbol. The gene symbol was used to find the overlap of the genes between the ABI Human Genome Survey Microarray version 2 array and the other arrays.

25

Validation using external data sets

We tested the performance of the gene signature developed in the FA versus FC classifier 1 based on Affymetrix HG-U133_plus2 array on the external data sets. Out of the 76 probe sets (66 unique genes) in the classifier 1 signature, 45 unique genes were
30 represented on the older HG-U133A array. In order to classify Webers data, we imported the 24 samples along with our 40 samples using only the approximately 22.200 probe sets that were shared between the HG-U133A and the HG-U133_plus2 arrays. The 24 samples were used (Weber *et al.* 2005) as an independent test set and we applied the classifier using the reduced signature of 45 genes. The same procedure
35 was used for the 12 sample validation data (Hinsch *et al.* 2009) employing the samples

as an independent test set implementing the gene signature from classifier 1 using the 53 genes out of the 66 genes in our classifier that were shared between the Affymetrix and the ABI platforms.

5 *Test of gene signatures across platforms*

In order to determine the overall performance of our model using the 76 probe set signature as compared to other published signatures and classifiers, we tested the two signatures reported by Weber et al. (3 and 80 genes) and one signature by Hinsch et al. (21 genes) using our data as the independent test set. In addition, we tested
10 signatures from other recent publications (Griffith *et al.* 2006; Foukakis *et al.* 2007; Prasad *et al.* 2008). These signatures counted a 5 gene signature (Foukakis *et al.* 2007), a 12 and 32 gene signature of benign versus malignant thyroid tumors from a meta-study (Griffith *et al.* 2006) and a 75 gene signature to discriminate benign from malignant thyroid nodules (Prasad *et al.* 2008). In this way a total of eight gene
15 signature was tested to assess their cross platform classification accuracy. All gene signatures were used in classifiers comparing the performance of DLDA, 1-NN, 3-NN and SVM, respectively, as implemented in the BRB-Array tools (Simon *et al.* 2007).

Immunohistochemistry

20 Resected tumors from thyroid glands were fixed by immersion with formalin. Paraffin sections were cut to a thickness of 4 um. The sections were placed in Target Retrieval Solution (DAKO, Denmark), and microwaved three times during 3 minutes to improve staining by antigen unmasking. After washing and quenching of endogenous
25 peroxidase, the sections were blocked and incubated for 1 hour at room temperature with antibodies against human Ki67 (Abcam), TOP2A (DAKO), NR4A1 (Lifespan biosciences), and NR4A3 (MBL). The labeling was visualized with Peroxidase labelled polymer conjugated to goat anti-rabbit (or -mouse) immunoglobulins (DAKO), followed by incubation with diaminobenzidine and counterstaining with hematoxylin.

30 Results

Clinical representation of patients and tumors in the study

We generated global expression profiles of 69 thyroid samples comprising 2 normal thyroid samples (NO), 9 nodular goiter (NG), 2 papillary carcinomas (PC), 4 anaplastic carcinomas (AC) as well as 52 samples of follicular neoplasia, 22 follicular adenomas
35 (FA), 18 follicular carcinomas (FC) and 12 fetal adenomas (FEA) samples (Table 1).

Furthermore, we collected 23 samples (14 PC and 9 NO) from external sites (E-GEOD-6004 and E-GEOD-7307) submitted to the expression profile repository at Array Express (<http://www.ebi.ac.uk/arrayexpress/>). The 69 samples were diagnosed by a pathologist to assign histopathological diagnosis (according to WHO classification).

5 Among the follicular neoplasia patients, women constituted the majority, 35 (67.3%), whereas only 17 (32.7%) of the patients were men. The median age of patients with follicular carcinoma was 65 years, which was higher than the median age of patients diagnosed with fetal adenoma (50 years) and follicular adenoma (54 years), respectively. The median size of the nodules was 5 cm in diameter (range 2 - 9.5 cm) in follicular carcinoma patients compared to 3.8 cm (range 2-8 cm) in the patient group with fetal adenoma nodules and 4.1 cm (range 2 - 11) for follicular adenoma patients (Table 1).

In order to obtain an overview of the molecular difference and similarities of the thyroid nodules, we compared the gene expression of all samples in a principal component analysis (PCA) using all transcripts (Figure 1). The three dimensional plot captures 36% of the variance measured across the 92 samples. The analysis showed that AC is clearly distinguishable from the other thyroid nodules, except for a few of the follicular carcinomas. NG is a specific entity and is easily distinguished from the remaining samples, although located next to normal thyroid and follicular adenoma samples. The FEAs are clustered together, but share similarity with FAs and FCs as expected. The FCs are the most heterogeneous group ranging from being in proximity to follicular adenomas and normal thyroid to the anaplastic carcinomas. Taken together the PCA based on global expression profiles is in agreement with the consensus, stating that AC may be safely diagnosed, whereas the major clinical challenge is to distinguish follicular carcinoma from adenoma. Moreover, the PCA indicates that FEAs are likely to represent a separate biological entity - distinguishable from FA and FC, although they have many shared features.

Differentially expressed transcripts

30 To achieve knowledge about the molecular perturbations leading to follicular neoplasia we compared the gene expression patterns of follicular carcinoma, adenoma and microfollicular nodules. Differentially expressed genes were identified by class comparison analysis as described in materials and methods. The comparative analyses FC vs FA, FEA vs FA and FC vs FEA resulted in the identification of 117, 240 and 512 differentially expressed probe sets, respectively (Tables 19, 20 and 21

respectively). Forty five probe sets were overlapping between FC versus FA and FC vs FEA, however, there was no overlap among the differentially expressed transcripts between FC and FA compared to FEA vs FA. Taken together, we inferred that the transcripts, which are changed in FC, are different from transcripts altered in FEA, indicating that the differences between FA and FC are likely to be cancer related. Moreover, the high number of selective FEA transcripts emphasizes the unique biological properties of this histopathological entity.

To provide an overview of the molecular function of the differentially expressed transcripts of the FC group compared to FA we categorized the encoded proteins as either DNA or RNA binding factors, extra-cellular matrix and adhesion and cytoskeletal components, proteins involved in metabolism or cell signalling, protein secretion, cell cycle regulation and apoptosis as well as DNA repair. Moreover, a few proteins or transcripts of unknown function were grouped together. The relative distribution of the functional groups is shown in Figure 2. Compared with the entire collection of probe sets on the chip (Jonson *et al.* 2007), where about 10% of the all proteins may be connected to cell cycle control and apoptosis, we noted that 21% of the differentially expressed mRNAs in FCs encoded proteins categorized under this group. Moreover, transcripts encoding factors involved in protein secretion or DNA repair were virtually absent and the number of mRNAs encoding DNA and RNA binding proteins was scarce. Transcripts encoding proteins involved in signalling and metabolism were represented similarly to the entire transcriptome.

Since cell cycle control and apoptosis are major cancer related pathways we examined these pathways in more detail. All redundant probe sets were excluded, and this left us with 101 unique differentially expressed factors. The function of every differentially expressed factor was derived from the comprehensive cDNA-supported gene and transcripts annotation Ace View (Thierry-Mieg & Thierry-Mieg 2006) (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html>) and substantiated by a search in PubMed and Google. This revealed a striking enrichment of transcripts encoding proteins involved in DNA replication and mitosis as well as apoptosis among the up-regulated and down-regulated mRNAs, respectively (Figure 3). In addition to proteins directly involved in cytokinesis, the increased level of RRM2 that previously has been implicated in carcinogenesis (Boukovinas *et al.* 2008; Souglakos *et al.* 2008) and TOP2A that may represent a possible treatment target were notable (Pritchard *et al.* 2008; O'Malley *et al.* 2009). Transcripts encoding the nuclear orphan receptors NR4A1 and NR4A3 were heavily down-regulated together with *JUN* and *FOSB* and the

two transcripts encoding growth inhibitory factors ERG2 and SDPR. Moreover, the down-regulation of the transcripts encoding the mitochondrial potassium voltage-gated channel KCNAB1 and the solute carrier organic anion transporter family 2A1 factors that both are implicated in Fas mediated apoptosis was observed.

5 We corroborated the correlation of the mitotic transcripts to cell division by Ki67 staining of a series of histological section of FA and FC and confirmed the up- and down-regulation of TOP2A, NR4A1 and NR4A3, respectively (Figure 4A). Nuclear Ki67 staining was mainly observed in the malignant epithelial follicular cells. A few scattered surrounding mesenchymal cells also stained positive in both adenoma and carcinoma.

10 In agreement with the microarray data shown in Figure 4B, FC exhibits a wide range of mitotic activity, but in general adenomas had fewer Ki67 positive cells. A similar pattern was observed for the TOP2A that was significantly elevated in the FC. Also in agreement with the microarray results, nuclear and cytoplasmic staining of NR4A1 and NR4A3 was reduced in carcinoma. Compared to the variable number of mitotic cells

15 the absence of NR4A1 and NR4A3 was observed in all carcinomas (Figure 4A). This led us to generate a hierarchical cluster of the transcripts in order to examine if this was a general pattern during transition from adenoma to carcinoma. As shown in Figure 4B, transcripts encoding apoptotic factors are consistently lost in carcinoma, whereas mitotic factors provide a gradient of increased expression among the malignant

20 nodules. To investigate if the loss of apoptotic factors was a primary sign of malignancy we compared the expression of the factors in normal thyroid gland with FA and FC as well as AC and PC. As shown in Figure 5, there is a clear tendency that the expression of the proteins is similar in adenoma and normal thyroid tissue, whereas the activity is lost in the carcinoma. Oppositely, mitotic factors are increased in the carcinoma group

25 in agreement with the expansion of malignant cells.

Generation of a robust classifier for follicular neoplasia

Since the analysis of differentially expressed transcripts provided a possible mechanism for cancer progression we aimed towards exploiting the results to generate

30 an accurate molecular classifier that could differentiate between benign and malignant follicular thyroid lesions. Three classes of follicular nodules were included in the analysis. Firstly, we focused on improving the ability to discriminate between FA and FC. This part is referred to as classifier 1. Secondly, we included the fetal adenomas and built a classifier that could distinguish between all three types of follicular lesions,

35 since the characteristics and relation to follicular carcinoma of these nodules is

incompletely understood. This is referred to as classifier 2. Furthermore, when introducing the FEA as independent test set in classifier 1 (FA versus FC), we observed that the FEA samples were mainly classified as FA samples (Figure 6B). Based on this observation, we decided to build a classifier that could distinguish
5 adenomas (FA and FEA combined) from carcinomas (FC). As a result, three different classifiers were generated and evaluated: classifier 1 (FA versus FC), classifier 2 (FA versus FEA versus FC) and classifier 3 (FA and FEA merged versus FC).

All classifiers were based on the support vector machine algorithm developed in our R script. We compared the accuracy of the prediction by using the different classification
10 methods implemented in BRB-Array tools. We selected 76 probe sets in order to discriminate between the two groups of 22 FA and 18 FC samples. Of the prediction algorithms that were tested, 3-Nearest Neighbour and support vector machine classification had the best overall performance in this study. The cross validation misclassification rate using the LOOCV procedure was ranging from 5% to 15% with
15 SVM out-performing all other classifiers. The performance of all six predictive algorithms is shown in Table 22.

Essentially, all the above described transcripts involved in cell division and apoptosis were included in the classifier. By applying cross-validation we ensured that the data used for evaluating the predictive accuracy of the classifier was distinct from the data
20 used to select the genes to include in and build the classifier. The overall accuracy and performance of the classifiers during cross validation are listed in Table 2. Classifier 1 achieved an accuracy of 95% during cross validation. Two out of the 40 samples were miss-classified, those counted one FA (FA18) and one FC (FC8) sample, respectively. Classifier 2 achieved a correct classification in 85% (44 out of 52 samples) during
25 LOOCV. The 8 misclassified samples were 2 FA, 2 FC and 4 FEA samples. Consistent with classifier 1, the samples FA18 and FC8 were misclassified. In classifier 3, FA and FEA merged versus FC, we observed a percent of accuracy of 90%. Fifty two (52) samples were used in this analysis, 5 of these received a wrong label by the classifier. Again, FA18 is classified as an FC sample, as are FA5 and FEA9. The samples FC8
30 and FC17 are classified as follicular adenomas.

The sensitivity for FC of classifier 1 was 0.94, and the specificity was 0.96, resulting in a NPV for FC of 0.96 (Table 2). The high performance of the classifier is also shown by the relative receiver operating characteristic (ROC) curve, where the area under curve AUC is 0.96 for classifier 1 (Figure 6C). Classifier 3, which was built to distinguish

between the two merged adenoma samples and the carcinomas, has a sensitivity of 0.89 for FC and a specificity of 0.91 for FA. When translating these numbers into the positive prediction values for FC, which are 0.94 for FC in classifier 1 compared to 0.84 for FC in classifier 2 and 3, this shows a better performance of classifier 1 to predict the follicular carcinomas correctly. To assess whether or not the classifiers predict more accurately than by chance (R Simon, 2007), we computed misclassification rates of 1000 random permutations in order to calculate a p-value of the global test that the classifier is picking up random noise in the data (Simon, R, 2007). The error rate estimate is statistical significant with p-values for the three classifiers of 0.01, 0.03 and 0.03, respectively (Table 2).

Predictive Probability and Principal Component Analysis of the classifiers

We used the trained SVM classifiers to derive the predictive probability of a sample being one or the other type using logit estimates (Platt JC 1999). A sample is classified as an FA sample if the predictive probability is above 0.5 and as an FC sample if the predictive probability is below 0.5 (Figure 6B). It should be noted, that a probability of 1.0 can be interpreted as the classifier being completely certain about its prediction, whereas a value of 0.5 reflects total uncertainty. The analysis showed that one sample (FC11) is close to the borderline ($p(\text{FA}) = 0.5$), although labeled correctly during LOO. One sample, FA18, is very similar to the FC samples having a probability of 0.96 of being an FC sample. This sample is misclassified in every analysis. Similarly, sample FC8 is classified as an FA sample, with a probability of belonging to the FA group of 0.83. When the FEA samples were included in the analysis as a test set, all but one FEA sample have a high probability of being FA samples (see samples labeled in green, Figure 6B). Interestingly, Principal Component Analysis (PCA) analysis (which does not use sample labels) of the 76 probe sets that constitute the gene signature of classifier 1, is in agreement with the classifier and showed full separation of the FA and FC classes, except for samples FA18 and FC8, which are located in the area between the FA and FC sample clusters (Figure 6A).

We also derived the prediction uncertainty (predictive probability) for assigning class labels to the FA, FC or FEA samples in classifier 2. The results are summarized in a triangular diagram of the probabilities (Figure 7B). Each vertex of the triangle represents a sub-class. Samples plotted close to a vertex have a high probability of belonging to this particular class, whereas samples that are plotted in the center are fully uncertain (Figure 7B). The eight samples that are misclassified are in concordance

with the cross validation error obtained during the training of the model; four FEA samples and two FA and FC, respectively. The plot shows that two thirds of the 12 samples are correctly classified as FEA. This is in concordance with the PCA analysis of the gene signature of classifier 2. Here we saw that the three classes are separable, although a few samples from each class are overlapping with other classes (Figure 7A). Driven by the observation that fetal adenoma samples resemble follicular adenomas according to classifier 1 (Figure 6B and Supplemental, Figure 1A) we constructed a third classifier, classifier 3, where the two adenoma subclasses were merged. We tested the probabilistic classifier's ability to distinguish between the merged adenoma group and the carcinomas. Results are shown in Table 2. Five samples were misclassified, three adenoma samples and two FC samples. In agreement with classifier 1 and 2, the sample, FA18 is placed as an FC sample (Supplemental, Figure 2B). According to the probabilistic classifier, ten out of the twelve fetal adenoma samples have a high probability of belonging to the follicular adenoma class (Supplemental, Figure 2B), which suggests that these two classes have a higher degree of similarity than follicular carcinoma and fetal adenoma. This is also supported by the differential expression analysis where we found 240 genes differentially expressed between follicular adenoma and fetal adenomas compared to 512 genes regulated between fetal adenomas and follicular carcinomas.

Test of classifier on external validation data

One down-side to microarray classifiers is that different studies analyzing the same outcome report different genes used in the classifier. Thus, we examined whether or not classifier 1 could accurately classify independent data. The validation data made publically available by Hinsch et al. was produced with oligonucleotide arrays from Applied Biosystems, and moreover, provided a means of validating the model cross platform. The data set consisted of 4 FA and 8 FC. We downloaded an expression matrix of preprocessed data and used gene symbols to match the genes in our model to the probe id on the ABI array. Of the 76 genes in our signature, 53 were represented on this array. Applying the SVM classifier on the data resulted in an accuracy of 83% (10/12), whereas the qLDA classifier had an accuracy of 92% (11/12) with sensitivity for FC of 1.0 validating the performance and robustness of the signature (Table 3, Table 18).

Moreover, we obtained raw data files of 12 FA and 12 FC analyzed with Affymetrix HG-U133A arrays, which were preprocessed and re-analyzed as described by Weber et al. Initially we reproduced the result of Weber et al., i.e. obtaining an accuracy of 96% (23/24) in discriminating 12 FA from 12 FC using linear discriminant analysis, based on 80 regulated genes (Table 3, Table 12). These samples were analyzed on an older generation of Affymetrix arrays, the HG-U133A array, and approximately 22.500 transcripts are shared between the two generations of arrays. Of the 76 probe sets in our classifier, 45 were represented on the older array and used in the analysis. When applying our SVM classifier on the Weber validation data, we obtained an accuracy of 92% (22/24) with a sensitivity of 0.83 and a specificity of 1.0 for FC (Tables 23 and 11), although this data set was preprocessed and normalized differently from our data. Multiple studies have shown that different normalization strategies show a great impact on data analysis and end results (Hoffmann *et al.* 2002; Ploner *et al.* 2005; Shedden *et al.* 2005) and the high accuracy on the validation data emphasizes the robustness of the classifier and signature. Although the reduced gene signature from classifier 1 performed well on the older generation of Affymetrix arrays we observed a decrease in performance if we substituted this reduced gene list into our classifier and applied it on the HG-U133_plus_2 array, resulting in an accuracy of 88% for SVM and 92% using the LDA algorithm (data not shown). This suggests that we would get even better classification of the validation data, had it been hybridized on the next generation arrays, since the subset from the HG-U133A array did not perform as well on the new array as the full set of 76 probe sets.

Cross platform and cross laboratory use of signatures

In order to further test the overall performance of the FA versus FC classifier we applied a selection of recent published gene signatures and classifiers on both our 40 follicular neoplasia samples and the 24 validation samples to see if our classifier gives comparable or better results. Aside from the 80 gene signature mentioned above we tested additional six gene signatures (Griffith *et al.* 2006; Foukakis *et al.* 2007; Prasad *et al.* 2008; Hinsch *et al.* 2009). The results for all the signatures are shown in Table 3 and in full for all employed algorithms in Tables 4 – 18 and Tables 22 and 23. Firstly, we applied the two signatures (3 and 80 genes) published by Weber et al. using the same model framework as in the previous analyses. The Weber 80 gene signature did not performed as good on our 40 samples (accuracy of 72%, (31/40), Tables 3, and 23) as our signature did on their data (accuracy of 92%, Table 3) although all genes from

Webers signature were represented on the array used in our analysis. When applying Webers optimized 3 gene signature on their own data, we obtained an accuracy of 83% (20/24) as compared to 43% (17/40) with our 40 samples (Tables 3, 4 and 12) showing that the 3 gene signature is too small to show any discriminate power on external data. On the contrary we got better results when applying the 5 gene signature published by Foukakis et al. on our 40 sample data set, namely an accuracy of 85% (34/40). Four of the misplaced samples are follicular carcinomas, which is equivalent to a sensitivity and specificity for classifying FC samples of 0.78 and 0.91 (Table 3). When applied on the 24 sample validation data set, the 5 gene signature had an accuracy of 71% (17/24), Table 13.

Also, we tested a 32 gene signature optimized to distinguish benign from malignant thyroid lesion, which was derived in a large meta-study taking the lessons from recent papers on thyroid microarray analysis into account (Griffith *et al.* 2006). Based on a ranking system giving higher weight to genes that are selected in three or more of the evaluated expression studies, a top twelve gene signature was devised (Griffith et al. 2006). For each signature we performed classification by applying both support vector machine (SVM) learning as well as other algorithms for comparison. The SVM model gave the worst outcome with the top 12 gene signature applied on our data of only 55% accuracy, i.e. sensitivity and specificity for FC of 0.44 and 0.64, respectively. Nearest centroid showed improved performance with 70% accuracy (Tables 3 and 6). Better results were obtained when applying the signature to the validation data set from Weber, showing 88% accuracy, reflecting a sensitivity and specificity of 0.83 and 0.92, respectively (Tables 3 and 14). Notably, this 12 gene signature performed as well as Weber's own 80 genes signature when applying the LDA algorithm resulting in an accuracy of 92% (22/24).

The poor results of the top 12 meta-genes on our data was improved somewhat when the full 32 gene signature was applied (Griffith *et al.* 2006), increasing the accuracy to 83% (33/40), see Tables 3 and 7. Lastly, we tested the performance of the SVM classification based on the 25 (21 annotated) gene signature published in Hinsch *et al.* on our 40 sample data set and the 24 sample set, which resulted in accuracies of 70% and 71% compared to 80% and 92% for the 75 gene signature published by Prasad et al. (Tables 3, 8, 10, 16 and 17).

In general we see that even if a signature showed good performance on one data set, it would perform poorly on the other data set (Table 3). Overall, classifier 1, built to classify follicular adenoma and carcinoma, shows the best cross-platform and cross-laboratory performance, both on the training set and on validation data sets (Hinsch *et al.* 2009; Weber *et al.* 2005) with a positive predictive value for malignancy for FC of 0.94 and 0.92 and 1.0, respectively (Tables 3, and 23).

Discussion

We showed that follicular carcinoma is characterized by increased levels of mRNAs encoding proteins involved in DNA replication and mitosis corresponding to increased numbers of dividing cells, as well as loss of transcripts encoding proteins involved in growth arrest and apoptosis. Taken together these aberrations may provide a minimal platform for malignant transformation (Evan & Vousden 2001b).

Poorly differentiated and invasive carcinomas are known to exhibit a high proliferative grading, and it has been debated whether the mitotic index was useful to diagnose follicular carcinoma (Perez-Montiel & Suster 2008; Ghossein 2009). In agreement with the clinical experience, we noted that cell-cycle mRNAs followed a gradient ranging from a few fold to more than 50 fold up-regulation, which may limit the isolated use of this parameter for diagnostic purposes. A number of the up-regulated transcripts including anillin (Hall *et al.* 2005), ARP 2/3 complex (Otsubo *et al.* 2004), abnormal spindle homolog (Ayllon & O'connor 2007; Lin *et al.* 2008), centromere protein F (Campone *et al.* 2008), KIF4A (Taniwaki *et al.* 2007), maternal embryonic leucine zipper kinase (Gray *et al.* 2005), NIMA-related kinase 2 (Hayward & Fry 2006), PDZ binding kinase, protein regulator of cytokinesis 1 (Boukarabila *et al.* 2009), regulator of chromosome condensation 2 (Stacey *et al.* 2008), encoded proteins that are directly involved in mitosis and several of them are over-expressed in other cancer. In particular transcripts encoding ribonucleotide reductase small subunit - RRM2 and topoisomerase 2 α - TOP2A, respectively, are intimately connected to cancer. RRM2 promotes invasion and metastasis of tumors and over-expression is associated with gemcitabine resistance (Boukovinas *et al.* 2008; Souglakos *et al.* 2008), whereas TOP2A is frequently amplified in breast cancer, where it is an independent predictor of survival and a marker for anthracycline based chemotherapy (Pritchard *et al.* 2008; O'Malley *et al.* 2009).

Apoptosis is important in both benign and malignant thyroid diseases (Mitsiades *et al.* 2000; Mitsiades *et al.* 2003; Chen *et al.* 2004; Mitsiades *et al.* 2006). Epithelial cells are under normal conditions strictly organized and detachment from the epithelial lining and basal membrane triggers apoptosis (Evan & Vousden 2001a). In this way proliferation and migration of preneoplastic cells are suppressed. By comparison of normal thyroid tissue and benign goiter and adenoma, we found that loss of apoptotic and growth arrest factors first occur during transformation to malignancy. Compared to papillary cancers that only exhibit a moderate decrease in apoptotic factors, down-regulation was marked in follicular cancers and anaplastic carcinomas. In contrast to the up-regulation of cell-cycle associated transcripts, loss of apoptic mRNAs was prominent in all samples implying that this event precedes proliferation. The coordinate down-regulation of *NR4A1* and *NR4A3* in combination with *JUN*, *FOSB* and *CITED2* is striking, since these factors previously have been shown to be part of a common proapoptotic- and cancer predisposing pathway (Mullican *et al.* 2007). The finding is moreover supported by two recent studies where *NR4A1* was found to be down-regulated in follicular carcinoma (Fryknas *et al.* 2006; Camacho *et al.* 2009). *NR4A1* and *NR4A3* also known as Nur77 and Nor-1, respectively, are homologous orphan nuclear receptors, that regulate the transcription of a common set of target genes (Li *et al.* 2006), and both have been described as homeostatic regulators of proliferation and apoptosis (Moll *et al.* 2006; Zhan *et al.* 2008). Whereas, *NR4A1* and *NR4A3* deficient mice, respectively, exhibit subtle phenotypes, it was recently shown that double knock-out quickly led to acute myeloid leukemia (AML). The mice exhibited abnormal expansion of hematopoietic stem cells (HSCs) and myeloid progenitors as well as decreased expression of the AP-1 transcription factors JunB and c-Jun and defective extrinsic apoptotic Fas-L and TRAIL signaling (Mullican *et al.* 2007). *NR4A1* and *NR4A3* translocate to mitochondria and stimulate release of cytochrome c in a BCL2 dependent manner. The observed down-regulation of mitochondrial ion channels could promote these processes, since they also participate in apoptosis (Yu & Choi 2000; Szabo *et al.* 2004).

Fetal adenoma has many similarities to follicular adenoma, but due to their morphological resemblance to fetal thyroid they have been described as separate follicular variant. As shown in the PCA of all transcripts both follicular carcinoma and adenoma are heterogenous, and this is also the case for FEA. FCs may roughly be distinguished from FEA by the same transcripts as those that differentiate FCs from

FAs and in this way classify FEA as adenoma. On the other hand a few hundred transcripts differ between FAs and FEAs supporting that FEA represents a distinct variant of the adenoma. We, however, found no evidence of any particular changes in fetal markers such as *PAX8*, *TTF-1*, *HEX* or markers of differentiated thyroid cells such as *DUOX*, *NIS*, *TPO* or *PDS* to support a fetal origin, so perhaps it is not justified to withhold the present nomenclature. Moreover, it should be noted that the difference among the transcriptomes of FA and FEA is merely a quantitative difference. We cannot detect mRNAs that are specific to either tumor, so taken together we propose that FEA are categorized as FAs until a unique difference in clinical outcome or biology has been demonstrated.

Since the global expression data identified a number of transcripts encoding factors intimately correlated to transformation we explored if it was possible to generate a robust diagnostic signature. We compared the performance of several algorithms (Table 2), and although the majority performed well, the support vector was efficient in all combinations. Classifier 1 that was designed to distinguish between FA and FC exhibited a sensitivity for FC of 0.94 and specificity of 0.96. Although with lower sensitivity and specificity, it was also possible to classify fetal adenoma supporting the unique nature of these tumors. One of the major challenges for microarray based diagnostics is to eliminate cross-platform variation and exploit public domain data in the development of robust signatures. Hence, we compared the efficacy of previously generated signatures on our material and used the publically available expression profiles as validation set. In agreement with previous experiences from analysis of breast cancer samples (Sotiriou & Pusztai 2009), there was limited overlap between the genetic signatures, that different research groups have employed for classification. Nevertheless, there was an encouraging agreement in their ability to predict the correct diagnosis (Prasad *et al.* 2008; Hinsch *et al.* 2009). Classifier 1 correctly determined the diagnosis of FA and FC in 92% of the tumors examined by two independent laboratories (Weber *et al.* 2005; Hinsch *et al.* 2009) and stands out as a very robust signature. In general classifiers consisting of few transcripts were less accurate, probably reflecting that small numbers of mRNAs in a classifier may be more sensitive to geographical- and platform variations (Table 3). We trust that the high accuracy of our signature may be related to improved mathematical tools and the fact that we have had the possibility to select more optimal probe sets, since the whole genome U133 2.0 array contains about double as many probe sets compared to previous generations of

arrays. The signature used for classification is moreover very similar to the set of differentially expressed mRNAs, which as described above, may reflect biological changes that are intimately connected to transformation.

- 5 In conclusion, we propose that down-regulation of factors involved in growth arrest and apoptosis may represent a decisive step in the pathogenesis of follicular carcinoma. The coordinate loss of NR4A1 and NR4A3 may play a central role in transformation since this pathway previously has been causally associated with malignancy. Finally, we propose that the described molecular pathways provide an accurate and robust
10 genetic signature that may be helpful when distinguishing follicular adenoma from carcinoma in clinical settings.

Reference List

- 15 Ayllon V & O'connor R 2007 PBK/TOPK promotes tumour cell proliferation through p38 MAPK activity and regulation of the DNA damage response. *Oncogene* 26 3451-3461.
- Barden CB, Shister KW, Zhu B, Guiter G, Greenblatt DY, Zeiger MA & Fahey TJ, III 2003 Classification of follicular thyroid tumors by molecular signature: results of gene profiling. *Clin.Cancer Res.* 9 1792-1800.
- 20 Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN & Edgar R 2009 NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37 D885-D890.
- Benjamini Y., Hochberg, and Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. 57, 289-300. 1995. *J. R. Statist. Soc. B.*
25 Ref Type: Generic
- Bolstad BM, Irizarry RA, Astrand M & Speed TP 2003 A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 19 185-193.
- 30 Boukarabila H, Saurin AJ, Batsche E, Mossadegh N, van LM, Otte AP, Pradel J, Muchardt C, Sieweke M & Duprez E 2009 The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev.* 23 1195-1206.
- Boukovinas I, Papadaki C, Mendez P, Taron M, Mavroudis D, Koutsopoulos A, Sanchez-Ronco M, Sanchez JJ, Trypaki M, Staphopoulos E, Georgoulis V, Rosell R & Souglakos J 2008 Tumor BRCA1, RRM1 and RRM2 mRNA expression levels and clinical response to first-line gemcitabine plus docetaxel in non-small-cell lung cancer patients. *PLoS.One.* 3 e3695.
- 35 Camacho CP, Latini FR, Oler G, Hojaij FC, Maciel RM, Riggins GJ & Cerutti JM 2009 Down-regulation of NR4A1 in follicular thyroid carcinomas is restored following lithium treatment. *Clin.Endocrinol.(Oxf)* 70 475-483.
- 40 Campone M, Campion L, Roche H, Gouraud W, Charbonnel C, Magrangeas F, Minvielle S, Geneve J, Martin AL, Bataille R & Jezequel P 2008 Prediction of

- metastatic relapse in node-positive breast cancer: establishment of a clinicogenomic model after FEC100 adjuvant regimen. *Breast Cancer Res. Treat.* 109 491-501.
- 5 Castro P, Sansonetty F, Soares P, Dias A & Sobrinho-Simoes M 2001 Fetal adenomas and minimally invasive follicular carcinomas of the thyroid frequently display a triploid or near triploid DNA pattern. *Virchows Arch.* 438 336-342.
- Chen S, Fazle Akbar SM, Zhen Z, Luo Y, Deng L, Huang H, Chen L & Li W 2004 Analysis of the expression of Fas, FasL and Bcl-2 in the pathogenesis of autoimmune thyroid disorders. *Cell Mol. Immunol.* 1 224-228.
- 10 Evan GI & Vousden KH 2001a Proliferation, cell cycle and apoptosis in cancer. *Nature* 411 342-348.
- Evan GI & Vousden KH 2001b Proliferation, cell cycle and apoptosis in cancer. *Nature* 411 342-348.
- Fagin JA & Mitsiades N 2008 Molecular pathology of thyroid cancer: diagnostic and clinical implications. *Best. Pract. Res. Clin. Endocrinol. Metab* 22 955-969.
- 15 Finley DJ, Zhu B, Barden CB & Fahey TJ, III 2004 Discrimination of benign and malignant thyroid nodules by molecular profiling. *Ann. Surg.* 240 425-436.
- Foukakis T, Gusnanto A, Au AY, Hoog A, Lui WO, Larsson C, Wallin G & Zedenius J 2007 A PCR-based expression signature of malignancy in follicular thyroid tumors. *Endocr. Relat Cancer* 14 381-391.
- 20 Fryknas M, Wickenberg-Bolin U, Goransson H, Gustafsson MG, Foukakis T, Lee JJ, Landegren U, Hoog A, Larsson C, Grimelius L, Wallin G, Pettersson U & Isaksson A 2006 Molecular markers for discrimination of benign and malignant follicular thyroid tumors. *Tumour. Biol.* 27 211-220.
- 25 Fujarewicz K, Jarzab M, Eszlinger M, Krohn K, Paschke R, Oczko-Wojciechowska M, Wiench M, Kukulska A, Jarzab B & Swierniak A 2007 A multi-gene approach to differentiate papillary thyroid carcinoma from benign lesions: gene selection using support vector machines with bootstrapping. *Endocr. Relat Cancer* 14 809-826.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY & Zhang J 30 2004 Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5 R80.
- Ghossein R 2009 Problems and controversies in the histopathology of thyroid carcinomas of follicular cell origin. *Arch. Pathol. Lab Med.* 133 683-691.
- 35 Gray D, Jubb AM, Hogue D, Dowd P, Kljavin N, Yi S, Bai W, Frantz G, Zhang Z, Koeppen H, de Sauvage FJ & Davis DP 2005 Maternal embryonic leucine zipper kinase/murine protein serine-threonine kinase 38 is a promising therapeutic target for multiple cancers. *Cancer Res.* 65 9751-9761.
- 40 Griffith OL, Melck A, Jones SJ & Wiseman SM 2006 Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *J. Clin. Oncol.* 24 5043-5051.
- 45 Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, Stacey SN, Skarphedinsson OB, Helgadottir H, Li W, Nagy R, Aguillo E, Faure E, Prats E, Saez B, Martinez M, Eyjolfsson GI, Bjornsdottir US, Holm H, Kristjansson K, Frigge ML, Kristvinsson H, Gulcher JR, Jonsson T, Rafnar T,

- Hjartarsson H, Mayordomo JI, de la CA, Hrafnkelsson J, Thorsteinsdottir U, Kong A & Stefansson K 2009 Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat.Genet.* 41 460-464.
- 5 Hall PA, Todd CB, Hyland PL, McDade SS, Grabsch H, Dattani M, Hillan KJ & Russell SE 2005 The septin-binding protein anillin is overexpressed in diverse human tumors. *Clin.Cancer Res.* 11 6780-6786.
- Hayward DG & Fry AM 2006 Nek2 kinase in chromosome instability and cancer. *Cancer Lett.* 237 155-166.
- Hegedus L 2004 Clinical practice. The thyroid nodule. *N.Engl.J.Med.* 351 1764-1771.
- 10 Hegedus L, Bonnema SJ & Bennedbaek FN 2003 Management of simple nodular goiter: current status and future perspectives. *Endocr.Rev.* 24 102-132.
- Hinsch N, Frank M, Doring C, Vorlander C & Hansmann ML 2009 QPRT: a potential marker for follicular thyroid carcinoma including minimal invasive variant; a gene expression, RNA and immunohistochemical study. *BMC.Cancer* 9 93.
- 15 Hoffmann R, Seidl T & Dugas M 2002 Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biol.* 3 RESEARCH0033.
- Jonson L, Vikesaa J, Krogh A, Nielsen LK, Hansen T, Borup R, Johnsen AH, Christiansen J & Nielsen FC 2007 Molecular composition of IMP1 ribonucleoprotein granules. *Mol.Cell Proteomics.* 6 798-811.
- 20 Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, Leskela S, Pita G, Milne R, Maravall J, Ramos I, Andia V, Rodriguez-Poyo P, Jara-Albarran A, Meoro A, del PC, Arribas L, Iglesias P, Caballero J, Serrano J, Pico A, Pomares F, Gimenez G, Lopez-Mondejar P, Castello R, Merante-Boschin I, Pelizzo MR, Mauricio D, Opocher G, Rodriguez-Antona C, Gonzalez-Neira A, Matias-Guiu X, Santisteban P & Robledo M 2009 The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS.Genet.* 5 e1000637.
- Li C & Wong WH 2001 Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc.Natl.Acad.Sci.U.S.A* 98 31-36.
- 30 Li QX, Ke N, Sundaram R & Wong-Staal F 2006 NR4A1, 2, 3--an orphan nuclear hormone receptor family involved in cell apoptosis and carcinogenesis. *Histol.Histopathol.* 21 533-540.
- Lin SY, Pan HW, Liu SH, Jeng YM, Hu FC, Peng SY, Lai PL & Hsu HC 2008 ASPM is a novel marker for vascular invasion, early recurrence, and poor prognosis of hepatocellular carcinoma. *Clin.Cancer Res.* 14 4814-4820.
- 35 Lubitz CC, Gallagher LA, Finley DJ, Zhu B & Fahey TJ, III 2005 Molecular analysis of minimally invasive follicular carcinomas by gene profiling. *Surgery* 138 1042-1048.
- Mazzanti C, Zeiger MA, Costouros NG, Umbricht C, Westra WH, Smith D, Somervell H, Bevilacqua G, Alexander HR & Libutti SK 2004 Using gene expression profiling to differentiate benign versus malignant thyroid tumors. *Cancer Res.* 64 2898-2903.
- 40 Mitsiades CS, Poulaki V, Fanourakis G, Sozopoulos E, McMillin D, Wen Z, Voutsinas G, Tseleni-Balafouta S & Mitsiades N 2006 Fas signaling in thyroid carcinomas is diverted from apoptosis to proliferation. *Clin.Cancer Res.* 12 3705-3712.

- Mitsiades CS, Poulaki V & Mitsiades N 2003 The role of apoptosis-inducing receptors of the tumor necrosis factor family in thyroid cancer. *J.Endocrinol.* 178 205-216.
- Mitsiades N, Poulaki V, Tseleni-Balafouta S, Koutras DA & Stamenkovic I 2000 Thyroid carcinoma cells are resistant to FAS-mediated apoptosis but sensitive to tumor necrosis factor-related apoptosis-inducing ligand. *Cancer Res.* 60 4122-4129.
- 5 Moll UM, Marchenko N & Zhang XK 2006 p53 and Nur77/p53 and Nur77/ TR3 - transcription factors that directly target mitochondria for cell death induction. *Oncogene* 25 4725-4743.
- Mullican SE, Zhang S, Konopleva M, Ruvolo V, Andreeff M, Milbrandt J & Conneely OM 2007 Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia. *Nat.Med.* 13 730-735.
- 10 Nikiforova MN, Lynch RA, Biddinger PW, Alexander EK, Dorn GW, Tallini G, Kroll TG & Nikiforov YE 2003 RAS point mutations and PAX8-PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. *J.Clin.Endocrinol.Metab* 88 2318-2326.
- 15 O'Malley FP, Chia S, Tu D, Shepherd LE, Levine MN, Bramwell VH, Andrulis IL & Pritchard KI 2009 Topoisomerase II alpha and responsiveness of breast cancer to adjuvant chemotherapy. *J.Natl.Cancer Inst.* 101 644-650.
- Otsubo T, Iwaya K, Mukai Y, Mizokami Y, Serizawa H, Matsuoka T & Mukai K 2004 Involvement of Arp2/3 complex in the process of colorectal carcinogenesis. *Mod.Pathol.* 17 461-467.
- 20 Perez-Montiel MD & Suster S 2008 The spectrum of histologic changes in thyroid hyperplasia: a clinicopathologic study of 300 cases. *Hum.Pathol.* 39 1080-1087.
- Platt JC 1999 Probabilities for SV machines. *Advances in Large Margin Classifiers* 61-74.
- 25 Ploner A, Miller LD, Hall P, Bergh J & Pawitan Y 2005 Correlation test to assess low-level processing of high-density oligonucleotide microarray data. *BMC.Bioinformatics.* 6 80.
- Prasad NB, Somervell H, Tufano RP, Dackiw AP, Marohn MR, Califano JA, Wang Y, Westra WH, Clark DP, Umbricht CB, Libutti SK & Zeiger MA 2008 Identification of genes differentially expressed in benign versus malignant thyroid tumors. *Clin.Cancer Res.* 14 3327-3337.
- 30 Pritchard KI, Messersmith H, Elavathil L, Trudeau M, O'Malley F & Dhesy-Thind B 2008 HER-2 and topoisomerase II as predictors of response to chemotherapy. *J.Clin.Oncol.* 26 736-744.
- 35 Ruggeri RM, Campenni A, Baldari S, Trimarchi F & Trovato M 2008 What is New on Thyroid Cancer Biomarkers. *Biomark.Insights.* 3 237-252.
- Shedden K, Chen W, Kuick R, Ghosh D, Macdonald J, Cho KR, Giordano TJ, Gruber SB, Fearon ER, Taylor JM & Hanash S 2005 Comparison of seven methods for producing Affymetrix expression scores based on False Discovery Rates in disease profiling data. *BMC.Bioinformatics.* 6 26.
- 40 Simon R 2006 A checklist for evaluating reports of expression profiling for treatment selection. *Clin.Adv.Hematol.Oncol.* 4 219-224.

- Simon, R., Lam A, Li, MC. Ngan M, Menenzes S, and Zhao Y. Analysis of Gene Expression Data using BRB-Array Tools. 2, 11-17. 2007.
Ref Type: Generic
- 5 Sotiriou C & Pusztai L 2009 Gene-expression signatures in breast cancer. *N.Engl.J.Med.* 360 790-800.
- 10 Souglakos J, Boukovinas I, Taron M, Mendez P, Mavroudis D, Tripaki M, Hatzidaki D, Koutsopoulos A, Stathopoulos E, Georgoulas V & Rosell R 2008 Ribonucleotide reductase subunits M1 and M2 mRNA expression levels and clinical outcome of lung adenocarcinoma patients treated with docetaxel/gemcitabine. *Br.J.Cancer* 98 1710-1715.
- 15 Stacey SN, Gudbjartsson DF, Sulem P, Bergthorsson JT, Kumar R, Thorleifsson G, Sigurdsson A, Jakobsdottir M, Sigurgeirsson B, Benediktsdottir KR, Thorisdottir K, Ragnarsson R, Scherer D, Rudnai P, Gurzau E, Koppova K, Hoim V, Botella-Estrada R, Soriano V, Juberias P, Grasa M, Carapeto FJ, Tabuenca P, Gilaberte Y, Gudmundsson J, Thorlacius S, Helgason A, Thorlacius T, Jonasdottir A, Blondal T, Gudjonsson SA, Jonsson GF, Saemundsdottir J, Kristjansson K, Bjornsdottir G, Sveinsdottir SG, Mouy M, Geller F, Nagore E, Mayordomo JI, Hansson J, Rafnar T, Kong A, Olafsson JH, Thorsteinsdottir U & Stefansson K 2008 Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat.Genet.* 40 1313-1318.
- 20 Szabo I, Adams C & Gulbins E 2004 Ion channels and membrane rafts in apoptosis. *Pflugers Arch.* 448 304-312.
- 25 Taniwaki M, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, Tsuchiya E, Kohno N, Nakamura Y & Daigo Y 2007 Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. *Clin.Cancer Res.* 13 6624-6631.
- Thierry-Mieg D & Thierry-Mieg J 2006 AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* 7 Suppl 1 S12-S14.
- Tusher VG, Tibshirani R & Chu G 2001 Significance analysis of microarrays applied to the ionizing radiation response. *Proc.Natl.Acad.Sci.U.S.A* 98 5116-5121.
- 30 Utiger RD 2005 The multiplicity of thyroid nodules and carcinomas. *N.Engl.J.Med.* 352 2376-2378.
- Vapnik, V. Statistical Learning Theory. 1998. New York, NY, USA, Wiley-Interscience.
Ref Type: Generic
- 35 Weber F, Shen L, Aldred MA, Morrison CD, Frilling A, Saji M, Schuppert F, Broelsch CE, Ringel MD & Eng C 2005 Genetic classification of benign and malignant thyroid follicular neoplasia based on a three-gene combination. *J.Clin.Endocrinol.Metab* 90 2512-2521.
- Yu SP & Choi DW 2000 Ions, cell volume, and apoptosis. *Proc.Natl.Acad.Sci.U.S.A* 97 9360-9362.
- 40 Zhan Y, Du X, Chen H, Liu J, Zhao B, Huang D, Li G, Xu Q, Zhang M, Weimer BC, Chen D, Cheng Z, Zhang L, Li Q, Li S, Zheng Z, Song S, Huang Y, Ye Z, Su W, Lin SC, Shen Y & Wu Q 2008 Cytosporone B is an agonist for nuclear orphan receptor Nur77. *Nat.Chem.Biol.* 4 548-556.

Classifier 1 – FA/FC signature (64 genes, 66 probes)

	Parametric p-value	t-value	Geom mean of intensities in class 1 (FA)	Geom mean of intensities in class 2 (FC)	Fold-change	Probe set	Gene symbol	Description
1	< 1e-07	-6,725	13,2732941	123,2595434	0,1076857	<u>201291 s at</u>	<u>TOP2A</u>	topoisomerase (DNA) II alpha 170kDa
2	1,00E-07	-6,435	7,356864	88,5105668	0,0831185	<u>209773 s at</u>	<u>RRM2</u>	ribonucleotide reductase M2 polypeptide
3	1,30E-06	-5,727	79,666928	206,7389928	0,3853503	<u>203228 at</u>	<u>PAFAH1B3</u>	platelet-activating factor acetylhydrolase, isoform Ib, gamma subunit 29kDa
4	1,60E-06	-5,672	76,7352849	240,1476255	0,3195338	<u>1554452 a at</u>	<u>HIG2</u>	hypoxia-inducible protein 2
5	1,80E-06	-5,633	338,3265618	933,2443549	0,3625273	<u>222581 at</u>	<u>XPR1</u>	xenotropic and polytropic retrovirus receptor
6	1,90E-06	-5,614	16,8166475	53,6311773	0,313561	<u>203276 at</u>	<u>LMNB1</u>	lamin B1
7	3,50E-06	-5,425	5,8103954	30,3321905	0,1915587	<u>219978 s at</u>	<u>NUSAP1</u>	nucleolar and spindle associated protein 1
8	5,00E-06	-5,31	7,4736769	21,9494226	0,3404954	<u>239005 at</u>	<u>NA</u>	narrow abdomen [<i>Drosophila melanogaster</i>]
9	6,90E-06	-5,209	11,985198	56,7328013	0,2112569	<u>207165 at</u>	<u>HMMR</u>	hyaluronan-mediated motility receptor (RHAMM)
10	7,20E-06	-5,198	7,5327771	47,7247823	0,1578379	<u>222608 s at</u>	<u>ANLN</u>	anillin, actin binding protein
11	7,80E-06	-5,168	6,6807033	17,4682922	0,3824474	<u>231849 at</u>	<u>KRT80</u>	keratin 80
12	1,09E-05	-5,065	678,7884139	1060,519623	0,6400527	<u>214501 s at</u>	<u>H2AFY</u>	H2A histone family, member Y
13	1,10E-05	-5,059	20,3416329	100,15642	0,2030986	<u>202705 at</u>	<u>CCNB2</u>	cyclin B2
14	1,49E-05	-4,964	10,7710795	58,0812286	0,1854485	<u>203358 s at</u>	<u>EZH2</u>	enhancer of zeste homolog 2 (<i>Drosophila</i>)
15	1,56E-05	-4,949	7,1941199	21,688959	0,331695	<u>223307 at</u>	<u>CDCA3</u>	cell division cycle associated 3
16	1,81E-05	-4,901	108,9311449	211,5712752	0,5148674	<u>201904 s at</u>	<u>CTDSPL</u>	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like
17	1,84E-05	-4,895	2107,568367	3378,629119	0,6237939	<u>200037 s at</u>	<u>CBX3</u>	chromobox homolog 3 (HP1 gamma homolog, <i>Drosophila</i>)
18	1,90E-05	-4,885	285,5752615	578,252361	0,4938592	<u>224578 at</u>	<u>RCC2</u>	regulator of chromosome condensation 2

19	2,30E-05	-4,823	187,4029713	329,3170409	0,5690655	<u>226914</u> at	<u>ARPC5L</u>	actin related protein 2/3 complex, subunit 5-like
20	2,77E-05	-4,764	8,310873	43,4782581	0,1911501	<u>207828</u> s at	<u>CENPF</u>	centromere protein F, 350/400ka (mitosin)
21	2,86E-05	-4,753	12,5099144	82,5832737	0,1514824	<u>242881</u> x at	<u>NA</u>	narrow abdomen [<i>Drosophila melanogaster</i>]
22	2,93E-05	-4,745	8,3187322	39,7012109	0,2095335	<u>204825</u> at	<u>MELK</u>	maternal embryonic leucine zipper kinase
23	3,76E-05	-4,665	4,4410754	23,4524485	0,1893651	<u>202095</u> s at	<u>BIRC5</u>	baculoviral IAP repeat-containing 5
24	3,91E-05	-4,652	12,1486769	60,2221476	0,201731	<u>202954</u> at	<u>UBE2C</u>	ubiquitin-conjugating enzyme E2C
25	3,92E-05	-4,651	12,3688236	52,1043989	0,2373854	<u>203755</u> at	<u>BUB1B</u>	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)
26	3,94E-05	-4,65	6,568002	26,3650633	0,2491176	<u>209642</u> at	<u>BUB1</u>	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)
27	3,95E-05	-4,649	27,8927489	91,667178	0,3042828	<u>231882</u> at	<u>LOC100131139</u>	similar to double homeobox A
28	4,49E-05	4,607	40,3895203	14,0364981	2,8774642	<u>240460</u> at	<u>NA</u>	narrow abdomen [<i>Drosophila melanogaster</i>]
29	4,28E-05	4,623	5003,203371	1726,804106	2,8973775	<u>200795</u> at	<u>SPARCL1</u>	SPARC-like 1 (hevin)
30	4,06E-05	4,64	2430,896642	1509,070433	1,610857	<u>224892</u> at	<u>PLDN</u>	pallidin homolog (mouse)
31	4,05E-05	4,641	2494,517949	696,1881082	3,5831091	<u>218723</u> s at	<u>C13orf15</u>	chromosome 13 open reading frame 15
32	3,91E-05	4,652	40,1737667	14,7456986	2,7244397	<u>219173</u> at	<u>MYO15B</u>	myosin XVB pseudogene
33	3,87E-05	4,655	135,5970343	30,343872	4,4686793	<u>214307</u> at	<u>HGD</u>	homogentisate 1,2-dioxygenase (homogentisate oxidase)
34	3,75E-05	4,666	131,1846036	19,8735503	6,6009647	<u>219877</u> at	<u>ZMAT4</u>	zinc finger, matrin type 4
35	3,69E-05	4,671	2835,810744	1658,376115	1,7099925	<u>204274</u> at	<u>EBAG9</u>	estrogen receptor binding site associated, antigen, 9
36	3,48E-05	4,69	3616,690923	2302,345823	1,5708721	<u>202371</u> at	<u>ICEA14</u>	transcription elongation factor A (SII)-like 4
37	3,37E-05	4,7	3380,87689	2015,143147	1,6777353	<u>212245</u> at	<u>MCFD2</u>	multiple coagulation factor deficiency 2
38	3,32E-05	4,705	784,6519792	320,92915	2,4449383	<u>225311</u> at	<u>IVD</u>	isovaleryl Coenzyme A dehydrogenase
39	2,88E-05	4,751	304,0536142	54,8002716	5,5483961	<u>218711</u> s at	<u>SDPR</u>	serum deprivation response (phosphatidylserine binding protein)
40	2,79E-05	4,761	67,9054071	44,717813	1,5185315	<u>213760</u> s at	<u>ZNF330</u>	zinc finger protein 330
41	2,70E-05	4,772	228,1179829	34,0074228	6,707888	<u>203029</u> s at	<u>PTPRN2</u>	protein tyrosine phosphatase, receptor type, N

65	6,00E-07	6,009	1118,792953	109,3897186	10,2275878	<u>205249</u> at	EGR2	early growth response 2 (Krox-20 homolog, Drosophila)
66	1,00E-07	6,46	287,1994253	27,9650994	10,2699233	<u>206209</u> s at	CA4	carbonic anhydrase IV

ITEMS

- 5 1. An mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21, and distinguishes between the classes thyroid follicular adenoma and thyroid follicular carcinoma, wherein said distinction is given as a prediction probability for said sample of belonging to either class, said probability being a number falling in the range of from 0 to 1.

10
- 15 2. An mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and distinguishes between the

20

25 classes thyroid follicular adenoma and thyroid follicular carcinoma, wherein said distinction is given as a prediction probability for said sample of belonging to either class, said probability being a number falling in the range of from 0 to 1.
- 30 3. The mRNA classifier according to item 1 wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3,

35

RCC2, CTDSP1, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, and ZNF330.

- 5 4. The mRNA classifier according to item 2 wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the group consisting of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1,

10 CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and SLCO2A1
- 15 5. An mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier comprises or consists of one or more mRNAs selected from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.
- 20 6. The mRNA classifier according to any of the preceding items, wherein said mRNA classifier comprises or consists of six or more mRNAs selected from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.
- 25 7. The mRNA classifier according to any of the preceding items, wherein said mRNA classifier comprises less than 120 mRNAs, such as less than 110 mRNAs, for example less than 100 mRNAs, such as less than 90 mRNAs, for example less than 80 mRNAs, such as less than 70 mRNAs, for example less than 60 mRNAs, such as less than 50 mRNAs, for example less than 40 mRNAs, such as less than 30 mRNAs, for example less than

30 20 mRNAs, such as less than 10 mRNAs.
8. The mRNA classifier according to any of the preceding items, wherein said mRNA classifier comprises less than 10 mRNAs.

- 5 9. The mRNA classifier according to any of the preceding items, wherein the sensitivity is at least 85%, such as at least 86%, for example at least 87%, such as at least 88%, for example at least 89%, such as at least 90%, for example at least 91%, such as at least 92%, for example at least 93%, such as at least 94%, for example at least 95%.
- 10 10. The mRNA classifier according to any of the preceding items, wherein the specificity is at least 85%, such as at least 86%, for example at least 87%, such as at least 88%, for example at least 89%, such as at least 90%, for example at least 91%, such as at least 92%, for example at least 93%, such as at least 94%, for example at least 95%.
- 15 11. The mRNA classifier according to any of the preceding items, wherein the prediction probability of a sample for belonging to a certain class is a number falling in the range of from 0 to 1, such as from 0.0 to 0.1, for example 0.1 to 0.2, such as 0.2 to 0.3, for example 0.3 to 0.4, such as 0.4 to 0.49, for example 0.5, such as 0.51 to 0.6, for example 0.6 to 0.7, such as 0.7 to 0.8, for example 0.8 to 0.9, such as 0.9 to 1.0.
- 20 12. The mRNA classifier according to any of the preceding items, wherein said classifier comprises or consists of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.
- 25 13. The mRNA classifier according to any of the preceding items, wherein an alteration of the expression profile of one or more of said mRNAs is associated with thyroid follicular carcinoma or thyroid follicular adenoma or fetal adenoma or thyroid follicular carcinoma and fetal adenoma.
- 30 14. The mRNA classifier according to any of the preceding items, wherein the up-regulation of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, and/or UBE2C, expression and/or down-regulation of AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1,

MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and/or SLCO2A1 expression is indicative of thyroid follicular carcinoma.

- 5 15. The mRNA classifier according to any of the preceding items, wherein the expression level of one or more mRNAs are determined by the microarray technique.
- 10 16. The mRNA classifier according to any of the preceding items, wherein the expression level of one or more mRNAs are determined by the quantitative polymerase chain reaction (QPCR) technique.
- 15 17. The mRNA classifier according to any of the preceding items, wherein the expression level of one or more mRNAs are determined by the northern blot technique.
18. The mRNA classifier according to any of the preceding items, wherein the expression level of one or more mRNAs are determined by Nuclease protection assay.
- 20 19. The mRNA classifier according to any of the preceding items, wherein the sample is extracted from an individual by fine-needle aspiration.
20. The mRNA classifier according to item 19, wherein the sample is extracted from an individual by single fine-needle aspiration.
- 25 21. The mRNA classifier according to item 19, wherein the sample is extracted from an individual by multiple fine-needle aspirations.
22. The mRNA classifier according to items 1-18, wherein the sample is extracted from an individual by coarse-needle aspiration.
- 30 23. The mRNA classifier according to item 1-18, wherein the sample is extracted from an individual by thyroid surgery.

24. The mRNA classifier according to item 1-18, wherein the sample is extracted from an individual by hemi-thyroidectomy.
25. The mRNA classifier according to items 1-18, wherein the sample is extracted from an individual by thyroid biopsy.
26. A model for predicting the diagnosis of an individual with a thyroid nodule, comprising
- i) providing a set of input data to the mRNA classifier according to any of the preceding items, and
 - ii) determining if said individual has a condition selected from the group of thyroid follicular adenoma, thyroid follicular carcinoma, and/or fetal adenoma.
27. The model according to item 26, wherein said input data comprises or consists of the mRNA expression profile of six or more of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1.
28. The model according to item 26, wherein said input data comprises or consists of the mRNA expression profile of six or more of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L,

CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, and ZNF330.

- 5 29. The model according to item 26, wherein said input data comprises or consists of the mRNA expression profile of six or more of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1, CCDC85A, CDH16, CITED2, 10 CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and SLCO2A1
- 15 30. The model according to item 26, wherein said input data comprises or consists of the mRNA expression profile of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.
- 20 31. A device for measuring the expression level of at least six mRNAs in a sample, wherein said device comprises or consists of probes selected from the groups disclosed in tables 19, 20, and 21.
- 25 32. A device according to item 31 for measuring the expression level of at least six mRNAs in a sample, wherein said device comprises or consists of probes selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, 30 H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, 35 SLC26A4, and SLCO2A1,

wherein said device is used for classifying a sample obtained from a thyroid nodule of an individual.

- 5 33. The device according to item 31, wherein said device comprises or consists of probes selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, 10 CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, and ZNF330.
- 15 34. The device according to item 31, wherein said device comprises or consists of probes selected from the group consisting of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1, CCDC85A, CDH16, CITED2, 20 CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and SLCO2A1
- 25 35. The device according to item 31, wherein said device comprises or consists of probes selected from the group consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.
- 30 36. The device according to items 31-35, wherein said device may be used for distinguishing between thyroid follicular adenoma and thyroid follicular carcinoma, and/or distinguishing thyroid follicular carcinoma and thyroid follicular adenoma merged with fetal adenoma.
- 35 37. The device according to any of items 31-35, wherein said device may be used to with the mRNA classifier according to any of items 1-25, to classify a sample into either of the classes of thyroid follicular adenoma, thyroid

follicular carcinoma, fetal adenoma or thyroid follicular adenoma merged with fetal adenoma.

5 38. The device according to any of items 31-35, wherein said device comprises less than 120 probes, such as less than 110 probes, for example less than 100 probes, such as less than 90 probes, for example less than 80 probes, such as less than 70 probes, for example less than 60 probes, such as less than 50 probes, for example less than 40 probes, such as less than 30 probes, for example less than 20 probes, such as less than 10 probes.

10

39. The device according to any of items 31-35, wherein said device comprises less than 10 probes.

15

40. The device according to any of items 31-35, wherein said device is a microarray chip.

41. The device according to item 40, wherein said device is a microarray chip comprising DNA probes.

20

42. The device according to item 41, wherein said device is a microarray chip comprising antisense mRNA probes.

25

43. The device according to any of items 31-35, wherein said device is a QPCR Microfluidic Card.

44. The device according to any of items 31-35, wherein said device comprises QPCR tubes, QPCR tubes in a strip or a QPCR plate.

30

45. The device according to any of items 31-35, wherein said device comprises probes on a solid support.

46. The device according to any of items 31-35, wherein said device comprises probes on at least one bead.

47. The device according to any of items 31-35, wherein said device comprises probes in liquid form in a tube.
- 5 48. A kit-of-parts comprising the device of any of items 31-35, and at least one additional component.
49. The kit according to item 48, wherein said additional component is means for extracting RNA, such as mRNA, from a sample.
- 10 50. The kit according to item 48, wherein said additional component is reagents for performing microarray analysis.
51. The kit according to item 48, wherein said additional component is reagents for performing QPCR analysis.
- 15 52. The kit according to item 48, wherein said additional component is the computer program product according to item 84.
53. The kit according to item 48, wherein said additional component is instructions for use of the device.
- 20 54. A method for determining the presence of a malignant and/or pre-malignant condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, selected from the groups disclosed in tables 19, 20, and 21.
- 25 55. A method for determining the presence of a malignant and/or pre-malignant condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, 30 CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3,
- 35

5 H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said expression level of said mRNAs is associated with thyroid follicular carcinoma or thyroid follicular adenoma merged with fetal adenoma.

10 56. A method for determining the presence of a benign condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the groups disclosed in tables 19, 20, and 21.

15 57. A method for determining the presence of a benign condition in a sample obtained from a thyroid nodule of an individual, said method comprising measuring the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the group consisting of FOSB, 20 LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, 25 CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, 30 MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said expression level of said mRNAs is associated with thyroid follicular adenoma.

58. A method for performing a diagnosis on an individual with a thyroid nodule, comprising the steps of:

- 5
- i) extracting RNA from a sample collected from the thyroid of an individual,
 - ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21.

59. A method for performing a diagnosis on an individual with a thyroid nodule, comprising the steps of:

- 10
- i) extracting RNA from a sample collected from the thyroid of an individual,
 - ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and
- 15
- iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma.
- 20
- 25

60. A method for diagnosing if an individual has, or is at risk of developing, follicular thyroid carcinoma and/or fetal adenoma, comprising the steps of:

- 30
- i) extracting RNA from a sample collected from the thyroid of an individual,
 - ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21,
- 35

wherein a predetermined mRNA expression profile of said mRNAs is indicative of the individual having, or being at risk of developing, follicular thyroid carcinoma and/or fetal adenoma.

- 5 61. A method for diagnosing if an individual has, or is at risk of developing, follicular thyroid carcinoma and/or fetal adenoma, comprising the steps of:
- i) extracting RNA from a sample collected from the thyroid of an individual,
- 10 ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
- 15 20 25 wherein a predetermined mRNA expression profile of said mRNAs is indicative of the individual having, or being at risk of developing, follicular thyroid carcinoma and/or fetal adenoma

62. The method according to item 60 or 61 wherein the six or more mRNAs are selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1,
- 30

MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, and ZNF330.

- 5 63. The method according to item 60 or 61 wherein the six or more mRNAs are selected from the group consisting of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, UBE2C, AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, 10 MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and SLCO2A1
- 15 64. The method according to item 60 or 61, wherein the mRNAs comprises or consists of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF.
- 20 65. The method according to item 60 or 61, wherein up-regulation of ANLN, ARPC5L, ASPM, BUB1B, CBX3, CCNB2, CDCA5, CENPF, CEP55, CKS2, CTD, H2A, KIF4A, MELK, NEK2, NUSAP1, PBK, PRC1, RCC2, RRM2, SAC3D1, TMPO, TOP2A, TPX2, and/or UBE2C, expression and/or down-regulation of AGTR1, CCDC85A, CDH16, CITED2, CTGF, CYR61, DLC1, DNASE1L3, DUSP14, EGR2, FOSB, JUN, KCNAB1, MAN1C1, MATN2, NR4A1, NR4A3, PLA2R1, PTPRN2, SDPR, SLC26A4, and/or SLCO2A1 expression is indicative of thyroid follicular carcinoma.
- 25 66. A method for determining the need for thyroidectomy in an individual presenting with a thyroid nodule, comprising the steps of:
- 30 i) extracting RNA from a sample collected from the thyroid of an individual,
- ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the groups disclosed I tables 19, 20, and 21,
- iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma, and

- iv) performing thyroidectomy on the individual only if the nodule is diagnosed as follicular thyroid carcinoma or fetal adenoma.

67. A method for determining the need for thyroidectomy in an individual presenting with a thyroid nodule, comprising the steps of:

5

- i) extracting RNA from a sample collected from the thyroid of an individual,
- ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group consisting of FOSB,
10 LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80,
15 PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO,
20 TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
- iii) determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma, and

15

20

25

performing thyroidectomy on the individual only if the nodule is diagnosed as follicular thyroid carcinoma or fetal adenoma

68. A method for performing thyroidectomy in a patient presenting with a thyroid nodule, comprising the steps of:

30

- i) extracting RNA from a sample collected from the thyroid of an individual,
- ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group consisting of FOSB,
35 LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5,

35

- 5 NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, 10 KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,
- iii) determining if said individual has a malignant or pre-malignant condition selected from follicular thyroid carcinoma or fetal adenoma.
- iv) performing thyroidectomy on the individual if the nodule is diagnosed as follicular thyroid carcinoma or fetal adenoma.
- 15 69. The method according to any of items 54 -68, wherein said method comprises obtaining prediction probabilities of between 0-1 for said sample.
- 20 70. The method according to any of items 54-68, wherein said method is used in combination with at least one additional diagnostic method.
- 25 71. The method according to item 70, wherein said at least one additional diagnostic method is selected from the group consisting of Scintillation counting, Blood sample analysis, Ultrasound imaging, Cytology, Histology and Assessment of risk factors.
- 30 72. The method according to item 70, wherein said at least one additional diagnostic method improves the sensitivity and/or specificity of the combined diagnostic outcome.
73. A method for expression profiling of a sample, comprising measuring six or more mRNAs selected from the groups disclosed in tables 19, 20, and 21, and correlating said expression profile to a clinical condition.

74. A method for expression profiling of a sample, comprising measuring six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, and correlating said expression profile to a clinical condition
75. The method according to items 73 -74, wherein said clinical condition is follicular thyroid carcinoma, follicular thyroid adenoma or fetal adenoma.
76. A method for determining the prognosis of an individual with a thyroid nodule, comprising the steps of
- i) extracting RNA from a sample collected from the thyroid of an individual,
 - ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the groups disclosed in tabled 19, 20, and 21,
 - iii) determining if said individual has a malignant or pre-malignant condition selected from follicular thyroid carcinoma or fetal adenoma
77. A method for determining the prognosis of an individual with a thyroid nodule, comprising the steps of
- i) extracting RNA from a sample collected from the thyroid of an individual,
 - ii) analysing the mRNA expression profile of the sample, comprising six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2,

5 ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5,
 NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2,
 BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B,
 BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80,
 PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1,
 C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12,
 SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1,
 EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55,
 CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO,
 10 TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN,
 KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1,

iii) determining if said individual has a malignant or pre-malignant
 condition selected from follicular thyroid carcinoma or fetal adenoma.

15 78. A system for determining the presence of a malignant and/or pre-malignant
 condition in a sample obtained from a thyroid nodule of an individual, said
 system comprising means for analysing the expression level of six or more
 mRNAs in said sample, wherein said mRNAs is selected from the groups
 disclosed in tables 19, 20, and 21, wherein said expression level of said
 20 mRNAs is associated with thyroid follicular carcinoma.

79. A system for determining the presence of a malignant and/or pre-malignant
 condition in a sample obtained from a thyroid nodule of an individual, said
 system comprising means for analysing the expression level of six or more
 25 mRNAs in said sample, wherein said mRNAs is selected from the group
 consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1,
 DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR,
 CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN,
 EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B,
 30 BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3,
 RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11,
 KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD,
 CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330,
 ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1,
 35 SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB,

JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said expression level of said mRNAs is associated with thyroid follicular carcinoma

- 5 80. A system for determining the presence of a benign condition in a sample obtained from a thyroid nodule of an individual, said system comprising means for analysing the expression level of six or more mRNAs in said sample, wherein said mRNAs are selected from the group consisting of
- 10 FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467,
- 15 MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, and SLCO2A1, wherein said
- 20 expression level of said mRNAs is associated with thyroid follicular adenoma.
81. A system for performing a diagnosis on an individual with a thyroid nodule, comprising:
- 25 i) means for analysing the mRNA expression profile of the thyroid nodule, comprising six or more mRNA selected from the disclosed in tables 19, 20, and 21, and
- ii) means for determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular
- 30 thyroid carcinoma and fetal adenoma
82. A system for performing a diagnosis on an individual with a thyroid nodule, comprising:
- i) means for analysing the mRNA expression profile of the thyroid nodule,
- 35 comprising six or more mRNA selected from the group of FOSB,

5 LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4,
MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2,
HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1,
UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1,
10 HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L,
CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB,
C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1,
SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5,
CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO,
15 TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1,
MATN2, NR4A3, SLC26A4, and SLCO2A1, and

ii) means for determining if said individual has a benign or a malignant/pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma.

15

83. A computer program product having a computer readable medium, said computer program product providing a system for predicting the diagnosis of an individual with a thyroid nodule, said computer program product comprising means for carrying out any of the steps of any of the systems according to any of items 78 to 82.

20

84. A system according to any of items 78 to 82, wherein the data is stored, such as stored in at least one database.

Table 1.

Clinical representation of data included in the training set. A total of 52 samples of follicular neoplasia were analyzed including 22 follicular adenoma (FA), 18 follicular carcinoma (FC) and 12 fetal adenoma (FEA) samples. The median age of patients with follicular carcinoma was 65 years, compared to 50 years for fetal carcinoma patients and 54 years for follicular adenoma patients; with median size of nodules being 5 cm (2; 9.5) for follicular carcinoma patients compared to 3.8 (2;8) cm for fetal adenoma nodules and 4.1(2;11) for follicular adenoma nodules.

Diagnosis	Age (years)	Sex (M/F)	Nodule size (cm; max diameter)	Relapse (Y/N)	Years from diagnosis
FC	76	M	4.5	N	4
FC	63	F	2.0	N	18
FC	90	M	6.0	Y*	3
FC	77	F	1.0	Y*	1 (deceased)
FC	67	M	8	Y	10
FC	52	F	5	Y	20
FC	61	M	6	Y	5
FC	56	M	6	Y	3
FC	71	F	5	Y	1
FC	88	F	2	Y	3
FC	24	F	2.5	N	1
FC	71	F	2.5	Y	7
FC	58	F	2.5	Y	13
FC	63	F	5	Y	5
FC	78	F	4	N	1
FC	74	M	9	N	2
FC	49	M	7	N	1
FC	61	F	9.5	N	1
FEA (n=12)**	50 (30;69)	6M/6F	3.8 (2;8)	-	-
FA (n=22)**	54 (31;77)	4M/18F	4.1 (2;11)	-	-

*) Metastasis found at the time of operation.

**) Data shown as the median value and the range of values are shown in parentheses.

Table 2 - Performance of the classifiers.

Performance of the three classifiers during leave-one-out training and cross validation. Classifier 1 is trained to predict and discriminate between follicular adenoma (FA) and follicular carcinomas (FC). Classifier 2 is trained to predict and discriminate between follicular adenoma (FA), follicular carcinoma (FC) and fetal adenoma (FEA). Classifier 3 is trained to discriminate between the merged adenomas versus the follicular carcinoma (FC). The table shows the overall classification performance of the three classifiers of 95%, 85% and 90% respectively. PPV is the positive predictive value, NPV is the negative predictive value. Accuracy, sensitivity, specificity, PPV and NPV is reported for each of the follicular subtypes taking part in each classification analysis during leave-one-out training and cross validation.

Model	Accuracy per Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Statistical significance(*)	Accuracy/error rate (%)
Classifier 1 class FA	95.5 (21/22)	95.5	94.4	95.5	94.4	P < 0.01	95% (2/40)
Classifier 1 class FC	94.4 (17/18)	94.4	95.5	94.4	95.5		
Classifier 2 class FA	90.9 (20/22)	90.9	86.7	83.3	92.9	P < 0.03	85% (8/52)
Classifier 2 Class FC	88.9 (16/18)	88.9	91.2	84.2	93.9		
Classifier 2 Class FEA	66.7 (8/12)	66.7	97.5	88.9	91.0		
Classifier 3 class FA + FEA	91.2 (31/34)	91.2	88.9	93.9	84.2	P < 0.03	90% (5/52)
Classifier 3 class FC	88.9 (16/18)	88.9	91.2	84.2	93.9		

(*) Statistical significance of error rate (1000 permutations).

Table 3 - Classifier performance of gene signatures.

Results obtained for classifier 1 on the internal 40 sample data set and the external data from Weber et al. (24 samples) and Hinsch et al. (12 samples). The SVM classifier is built to distinguish between FA and FC samples. Accuracy and sensitivity for FC is shown.

5

Table name	Gene signature used	Data set	Accuracy	Sensitivity (FC)
Table 2, 3, 2 supp.	76 genes (76 probes) classifier 1	40 samples	95%	0.94
Tables 3, 10 supp.	76 genes (45 probes) classifier 1	24 samples	92%	0.83
Tables 3, 18 supp.	76 genes (53 probes) classifier 1	12 samples	83% (92% qLDA)	0.88 (1.0)
Tables 3, 3 supp.	80 genes (96PS) Weber et al.	40 samples	72%	0.67
Tables 3, 11 supp.	80 genes (96PS) Weber et al.	24 samples	92%	0.91
Tables 3, 4 supp.	3 genes (5PS) Weber et al.	40 samples	43%	0.28
Tables 3, 12 supp.	3 genes (5PS) Weber et al.	24 samples	83%	0.92
Tables 3, 5 supp.	5 genes (6PS) Foukakis et al.	40 samples	85%	0.78
Tables 3, 13 supp.	5 genes (6PS) Foukakis et al.	24 samples	71%	0.67
Tables 3, 6 supp.	12 genes (28 PS) Griffith et al.	40 samples	55%	0.44
Tables 3, 14 supp.	12 genes (28 PS) Griffith et al.	24 samples	88%	0.83
Tables 3, 7 supp.	32 genes (32 PS) Griffith et al.	40 samples	83%	0.78
Tables 3, 15 supp.	32 genes (26 PS) Griffith et al.	24 samples	75%	0.75
Tables 3, 8 supp.	21 genes (32 PS) Hinsch et al.	40 samples	70%	0.67
Tables 3, 16 supp.	21 genes (27 PS) Hinsch et al.	24 samples	71%	0.58
Tables 3, 9 supp.	75 genes (158 PS) Prasad et al.	40 samples	80%	0.67
Tables 3, 17 supp.	75 genes (93 PS) Prasad et al.	24 samples	92%	0.83

Table 4. Classification results obtained using the 3 gene signature published by (Weber et. al) and the 40 FA and FC samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	54.5 (21/22)	54.5	61.1	63.2	52.4	57.5% (17/40)
	FC	61.1 (11/18)	61.1	54.5	52.4	63.2	
Diagonal Linear Discriminant Analysis	FA	54.5 (10/22)	54.5	44.4	54.5	44.4	50% (20/40)
	FC	44.4 (10/18)	44.4	54.5	44.4	54.5	
1-Nearest Neighbor	FA	63.6 (14/22)	63.6	66.7	70.0	60.0	65% (14/40)
	FC	66.7 (12/18)	66.7	63.6	60.0	70.0	
3-Nearest Neighbor	FA	77.3 (17/22)	77.3	61.1	70.8	68.8	70% (12/40)
	FC	61.1 (11/18)	61.1	77.3	68.8	70.8	
Nearest Centroid	FA	54.5 (12/22)	54.5	66.7	66.7	54.5	60% (16/40)
	FC	66.7 (12/18)	66.7	54.5	54.5	66.7	
Support Vector Machines	FA	54.5 (12/22)	54.5	27.8	48.0	33.3	42.5% (23/40)
	FC	27.8 (5/18)	27.8	54.5	33.3	48.0	

Table 5. Classification results obtained using the 5 gene signature published by (Foukakis et. al) and the 40 FA and FC samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	86.4 (19/22)	86.4	72.2	79.2	81.2	80% (8/40)
	FC	72.2 (13/18)	72.2	86.4	81.2	79.2	
Diagonal Linear Discriminant Analysis	FA	90.9 (20/22)	90.9	77.8	83.3	87.5	85% (6/40)
	FC	77.8 (14/18)	77.8	90.9	87.5	83.3	
1-Nearest Neighbor	FA	59.1 (20/22)	59.1	61.1	65.0	55.0	60% (16/40)
	FC	61.1 (14/18)	61.1	59.1	55.0	65.0	
3-Nearest Neighbor	FA	72.7 (16/22)	72.7	61.1	69.6	64.7	68% (13/40)
	FC	61.1 (11/18)	61.1	72.7	64.7	69.6	
Nearest Centroid	FA	90.9 (20/22)	90.9	77.8	83.3	87.5	85% (6/40)
	FC	77.8 (14/18)	77.8	90.9	87.5	83.3	
Support Vector Machines	FA	90.9 (20/22)	90.9	77.8	83.3	87.5	85% (6/40)
	FC	77.8 (14/18)	77.8	90.9	87.5	83.3	

Table 6. Classification results obtained using the 12 gene signature published by Griffith et. al. and the 40 FA and FC samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	81.8% (18/22)	81.8	44.4	64.3	66.7	65% (14/40)
	FC	44.4% (8/18)	44.4	81.8	66.7	64.3	
Diagonal Linear Discriminant Analysis	FA	81.8% (18/22)	81.8	44.4	64.3	66.7	65% (14/40)
	FC	44.4% (8/18)	44.4	81.8	66.7	64.3	
1-Nearest Neighbor	FA	81.8% (18/22)	81.8	44.4	64.3	66.7	65% (14/40)
	FC	44.4% (8/18)	44.4	81.8	66.7	64.3	
3-Nearest Neighbor	FA	90.9% (20/22)	90.9	27.8	60.6	71.4	62.5% (15/40)
	FC	27.8 (5/18)	27.8	90.9	71.4	60.6	
Nearest Centroid	FA	81.8 (18/22)	81.8	55.6	69.2	71.4	70% (12/40)
	FC	55.6 (10/18)	55.6	81.8	71.4	69.2	
Support Vector Machines	FA	63.6% (14/22)	63.6	44.4	58.3	50.0	55% (18/40)
	FC	44.4% (8/18)	44.4	63.6	50.0	58.3	

Table 7. Classification results obtained using the 32 gene signature published by Griffith et al. and the 40 FA and FC samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	81.8 (18/22)	81.8	66.7	75.0	75.0	75% (10/40)
	FC	66.7 (12/18)	66.7	81.8	75.0	75.0	
Diagonal Linear Discriminant Analysis	FA	77.3 (17/22)	77.3	66.7	73.9	70.6	72.5% (11/40)
	FC	66.7 (12/18)	66.7	77.3	70.6	73.9	
1-Nearest Neighbor	FA	77.3 (17/22)	77.3	44.4	63.0	61.5	62.5% (15/40)
	FC	44.4 (8/18)	44.4	77.3	61.5	63.0	
3-Nearest Neighbor	FA	72.7 (16/22)	72.7	50.0	64.0	60.0	62.5% (15/40)
	FC	50.0 (9/18)	50.0	72.7	60.0	64.0	
Nearest Centroid	FA	77.3 (17/22)	77.3	66.7	73.9	70.6	72.5% (11/40)
	FC	66.7 (12/18)	66.7	77.3	70.6	73.9	
Support Vector Machines	FA	86.4 (19/22)	86.4	77.8	82.6	82.4	82.5% (7/40)
	FC	77.8 (14/18)	77.8	86.4	82.4	82.6	

Table 8. Classification results obtained using the 21 gene signature published by Hinsch et al. and the 40 FA and FC samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	72.7% (16/22)	72.7	55.6	66.7	62.5	65% (14/40)
	FC	55.6% (10/18)	55.6	72.7	62.5	66.7	
Diagonal Linear Discriminant Analysis	FA	77.3% (17/22)	77.3	50.0	65.4	64.3	65% (14/40)
	FC	50.0% (9/18)	50.0	77.3	64.3	65.4	
1-Nearest Neighbor	FA	72.7% (16/22)	72.7	50.0	64.0	60.0	62.5% (15/40)
	FC	50.0% (9/18)	50.0	72.7	60.0	64.0	
3-Nearest Neighbor	FA	72.7% (16/22)	72.7	38.9	59.3	53.8	57.5% (17/40)
	FC	38.9 (7/18)	38.9	72.7	53.8	59.3	
Nearest Centroid	FA	72.7 (16/22)	72.7	55.6	57.1	0.5	55% (18/40)
	FC	33.3 (6/18)	33.3	72.7	0.5	57.1	
Support Vector Machines	FA	72.7% (16/22)	72.7	66.7	72.7	66.7	70% (12/40)
	FC	66.7% (12/18)	66.7	72.7	66.7	72.7	

Table 9. Classification results obtained using the 75 gene signature (represented by 158 probe sets on the HG-U133A array) published by Prasad et al., and the 40 FA and FC samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	86.4 (19/22)	86.4	77.8	82.6	82.4	82.5% (7/40)
	FC	77.8 (14/18)	77.8	86.4	82.4	82.6	
Diagonal Linear Discriminant Analysis	FA	86.4 (19/22)	86.4	77.8	82.6	82.4	82.5% (7/40)
	FC	77.8 (14/18)	77.8	86.4	82.4	82.6	
1-Nearest Neighbor	FA	90.9 (20/22)	90.9	72.2	80.0	86.7	82.5% (7/40)
	FC	72.2 (13/18)	72.2	90.9	86.7	80.0	
3-Nearest Neighbor	FA	95.5 (21/22)	95.5	61.6	75.0	100	80.0% (8/40)
	FC	61.6 (11/18)	61.6	95.5	91.7	91.7	
Nearest Centroid	FA	86.4 (19/22)	86.4	72.2	79.2	81.2	80% (8/40)
	FC	72.2 (13/18)	72.2	66.7	81.2	79.2	
Support Vector Machines	FA	90.9 (20/22)	90.9	66.7	76.9	85.7	80% (8/40)
	FC	66.7 (12/18)	66.7	90.9	85.7	76.9	

Table 10. Classification results obtained using the 76 gene signature formulated in classifier 1 and the 24 FA and FC (weber et al.) samples as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	91.7 (11/12)	91.7	83.3	84.6	90.9	88% (3/24)
	FC	83.3 (10/12)	83.3	91.7	90.9	84.6	
Diagonal Linear Discriminant Analysis	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	
1-Nearest Neighbor	FA	91.7 (11/12)	91.7	75.0	78.6	0.90	83% (4/24)
	FC	75.0 (9/12)	75.0	91.7	0.90	78.6	
3-Nearest Neighbor	FA	100 (12/12)	100	66.7	75.0	100	83% (4/24)
	FC	91.7 (8/12)	66.7	100	100	75.0	
Nearest Centroid	FA	91.7 (11/12)	91.7	75.0	78.6	90.0	83% (4/24)
	FC	75.0 (9/12)	75.0	91.7	90.0	78.6	
Support Vector Machines	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	

Table 11. Classification results obtained using the 80 gene (96 probe set) signature published by (Weber et. al) and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	100 (12/12) 91.7	100	91.7	92.3	100	96% (23/24)
	FC	(11/12)	91.7	100	100	92.3	
Diagonal Linear Discriminant Analysis	FA	100 (12/12) 91.7	100	91.7	92.3	100	96% (23/24)
	FC	(11/12)	91.7	100	100	92.3	
1-Nearest Neighbor	FA	91.7 (11/12) 91.7	91.7	91.7	91.7	91.7	92% (22/24)
	FC	(11/12)	91.7	91.7	91.7	91.7	
3-Nearest Neighbor	FA	100 (12/12) 91.7	100	91.7	92.3	100	96% (23/24)
	FC	(11/12)	91.7	100	100	92.3	
Nearest Centroid	FA	100 (12/12) 91.7	100	91.7	92.3	100	96% (23/24)
	FC	(11/12)	91.7	100	100	92.3	
Support Vector Machines	FA	91.7 (11/12) 91.7	91.7	91.7	91.7	91.7	92% (22/24)
	FC	(11/12)	91.7	91.7	91.7	91.7	

Table 12. Classification results obtained using the 3 gene signature published by (Weber et. al) and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	75.0 (9/12)	75.0	100	100	80.0	88% (3/24)
	FC	100 (12/12)	100	75.0	80.0	100	
Diagonal Linear Discriminant Analysis	FA	75.0 (9/12)	75.0	100	100	80.0	88% (3/24)
	FC	100 (12/12)	100	75.0	80.0	100	
1-Nearest Neighbor	FA	83.3 (10/12)	83.3	91.7	90.9	84.6	88% (3/24)
	FC	91.7 (11/12)	91.7	83.3	84.6	90.9	
3-Nearest Neighbor	FA	91.7 (11/12)	91.7	100	100	92.3	96% (1/24)
	FC	100 (12/12)	100	91.7	92.3	100	
Nearest Centroid	FA	75.0 (9/12)	75.0	100	100	80.0	88% (3/24)
	FC	100 (12/12)	100	75.0	80.0	100	
Support Vector Machines	FA	75.0 (9/12)	75.0	91.7	90.0	78.6	83% (4/24)
	FC	91.7 (11/12)	91.7	75.0	78.6	90.0	

Table 13. Classification results obtained using the 5 gene signature published by (Foukakis et al.) and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	91.7 (11/12)	91.7	58.3	68.8	87.5	75% (6/24)
	FC	58.3 (7/12)	58.3	91.7	87.5	68.8	
Diagonal Linear Discriminant Analysis	FA	83.3 (10/12)	83.3	66.7	71.4	80.0	75% (6/24)
	FC	66.7 (8/12)	66.7	83.3	80.0	71.4	
1-Nearest Neighbor	FA	66.7 (8/12)	66.7	91.7	88.9	73.3	79% (5/24)
	FC	91.7 (11/12)	91.7	66.7	73.3	88.9	
3-Nearest Neighbor	FA	83.3 (10/12)	83.3	75.0	76.9	81.8	79% (5/24)
	FC	75.0 (9/12)	75.0	83.3	81.8	76.9hb	
Nearest Centroid	FA	83.3 (10/12)	83.3	75.0	76.9	81.8	790% (5/24)
	FC	75.0 (9/12)	75.0	83.3	81.8	76.9	
Support Vector Machines	FA	75.0 (9/12)	75.0	66.7	69.2	72.7	71% (7/24)
	FC	66.7 (8/12)	66.7	75.0	72.7	69.2	

Table 14. Classification results obtained using the 12 gene signature published by Griffith et al. and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	100 (12/12)	100	66.7	75.0	100	83% (4/24)
	FC	66.7 (8/12)	66.7	100	100	75.0	
Diagonal Linear Discriminant Analysis	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	
1-Nearest Neighbor	FA	100 (12/12)	100	75.0	80.0	100	88% (3/24)
	FC	75 (9/12)	75.0	100	100	80.0	
3-Nearest Neighbor	FA	91.7 (11/12)	91.7	50.0	64.7	85.7	71% (7/24)
	FC	50 (6/12)	50.0	91.7	85.7	64.7	
Nearest Centroid	FA	100 (12/12)	100	58.3	70.6	100	79% (5/24)
	FC	58.3 (7/12)	58.3	100	100	70.6	
Support Vector Machines	FA	91.7 (11/12)	91.7	83.3	84.6	90.9	88% (3/24)
	FC	83.3 (10/12)	83.3	91.7	90.9	84.6	

Table 15. Classification results obtained using the 32 gene signature (represented by 26 probe sets on the HG-U133A array) published by Griffith et al., and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	
Diagonal Linear Discriminant Analysis	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	
1-Nearest Neighbor	FA	83.3 (10/12)	83.3	75.0	76.9	81.8	79% (5/24)
	FC	75.0 (9/12)	75.0	83.3	81.8	76.9	
3-Nearest Neighbor	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	
Nearest Centroid	FA	100 (12/12)	100	83.3	85.7	100	92% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	
Support Vector Machines	FA	75 (9/12)	75	75	75	75	75% (6/24)
	FC	75 (9/12)	75	75	75	75	

Table 16. Classification results obtained using the 21 gene signature published by Hinsch et al. and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

5

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	66.7 (8/12)	100	83.3	80.0	71.4	75% (6/24)
	FC	83.3 (10/12)	83.3	100	71.4	80.0	
Diagonal Linear Discriminant Analysis	FA	58.3 (7/12)	58.3	75.0	70.0	64.3	66.7% (8/24)
	FC	75.0 (9/12)	75.0	58.3	64.3	70.0	
1-Nearest Neighbor	FA	83.3 (10/12)	83.3	58.3	66.7	77.8	70.8% (7/24)
	FC	58.3 (7/12)	58.3	83.3	77.8	66.7	
3-Nearest Neighbor	FA	83.3 (10/12)	83.3	58.3	66.7	77.8	70.8% (7/24)
	FC	58.3 (7/12)	58.3	83.3	77.8	66.7	
Nearest Centroid	FA	58.3 (7/12)	58.3	83.3	77.8	66.7	70.8% (7/24)
	FC	83.3 (10/12)	83.3	58.3	66.7	77.8	
Support Vector Machines	FA	83.3 (10/12)	83.3	58.3	66.7	77.8	70.8% (7/24)
	FC	58.3 (7/12)	58.3	83.3	77.8	66.7	

Table 17. Classification results obtained using the 75 gene signature (represented by 93 probe sets on the HG-U133A array) published by Prasad et al., and the 24 FA and FC samples (Weber et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	83.3 (10/12)	83.3	83.3	83.3	83.3	83.3% (4/24)
	FC	83.3 (10/12)	83.3	83.3	83.3	83.3	
Diagonal Linear Discriminant Analysis	FA	83.3 (10/12)	83.3	75.0	76.9	81.8	79.2% (5/24)
	FC	75.0 (9/12)	75.0	83.3	81.8	76.9	
1-Nearest Neighbor	FA	91.7 (11/12)	91.7	83.3	84.6	90.9	87.5% (3/24)
	FC	83.3 (10/12)	83.3	91.7	90.9	84.6	
3-Nearest Neighbor	FA	100 (12/12)	100	66.7	75.0	100	83.3% (4/24)
	FC	66.7 (8/12)	66.7	100	100	75.0	
Nearest Centroid	FA	83.3 (10/12)	83.3	83.3	83.3	83.3	83.3% (4/24)
	FC	83.3 (10/12)	83.3	83.3	83.3	83.3	
Support Vector Machines	FA	100 (12/12)	100	83.3	85.7	100	91.7% (2/24)
	FC	83.3 (10/12)	83.3	100	100	85.7	

Table 18. Classification results obtained using the 76 gene signature formulated in classifier 1 (represented by 53 probes on the ABI Human Genome Survey Microarray version 2 array) and the 12 (4 FA and 8 FC samples (Hinsch et al.) as test set. The table shows the overall classification results when applying different classification algorithms. Results achieved with the Support vector machine (SVM) classifier are highlighted.

Model	Class	Accuracy Class (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy /error rate of classifier
Compound Covariate Predictor	FA	75.0 (3/4)	75.0	100	100	88.9	92% (1/12)
	FC	100 (8/8)	100	75.0	88.9	100	
Diagonal Linear Discriminant Analysis	FA	75.0 (3/4)	75.0	100	100	88.9	92% (1/12)
	FC	100 (8/8)	100	75.0	88.9	100	
1-Nearest Neighbor	FA	50.0 (2/4)	50.0	75.0	50.0	75.0	67% (4/12)
	FC	75.0 (6/8)	75.0	50.0	75.0	50.0	
3-Nearest Neighbor	FA	75.0 (3/4)	75.0	75.0	60.0	85.7	75% (3/12)
	FC	75.0 (6/8)	75.0	75.0	85.7	60.0	
Nearest Centroid	FA	75.0 (3/4)	75.0	87.5	75.0	87.5	83% (2/12)
	FC	87.5 (7/8)	87.5	75.0	87.5	75.0	
Support Vector Machines	FA	75.0 (3/4)	75.0	87.5	75.0	87.5	83% (2/12)
	FC	87.5 (7/8)	87.5	75.0	87.5	75.0	

Table 19 (FC_vs_FA_117) lists the 117 probe sets found to be differentially expressed in the class comparison between FA and FC.

Gene Ontology annotation is shown as is the result of a multiple testing correction algorithm. The base line is set to the FA class and the experiment is set to the FC class. Raw p is the p-value of Student's t-test. Genes are filtered to have a Benjamini Hochberg corrected p values below 0.05, an absolute fold change above 1.5 as well as an absolute change in expression levels of more than 100.

Probe set	symbol	Gene	Accession	Baseline mean	Experiment mean	Fold change	Raw p-value	Hochberg
1553243_at	ITIH5	inter-alpha (globulin) inhibitor H5	NM_032817	194,04	67,99	-2,85	7,77E-07	0,042447775
1554452_at	HIG2	hypoxia-inducible protein 2	BC001863	82,9	336,41	4,06	1,61E-06	0,08765308
1557107_at	LOC286002	hypothetical protein LOC286002	BC037315	1926	801,69	-2,4	1,73E-05	0,941639663
1557965_at	MTERFD2	MTERF domain containing 2	AL566167	647,48	373,27	-1,73	0,000126	1
1570289_at	LOC646736	similar to Alu subfamily SX sequence contamination warning entry	BC017935	271,45	107,69	-2,52	4,95E-05	1
160020_at	MMP14	matrix metalloproteinase 14 (membrane-inserted) chromobox homolog 3	Z48481	52,91	153,59	2,9	4,96E-05	1
200037_s_at	CBX3	(HP1 gamma homolog, Drosophila) /// similar to chromobox homolog 3	NM_016587	2150,72	3615,96	1,68	1,84E-05	1
200795_at	SPARCL1	SPARC-like 1 (mast9, hevjin) heterogeneous nuclear	NM_004684	5335,45	2668,11	-2	4,28E-05	1
201277_s_at	HNRPAB	ribonucleoprotein A/B	NM_004499	1498,85	2483,24	1,66	0,00013	1
201289_at	CYR61	cysteine-rich, angiogenic inducer, 61	NM_001554	10775,4	2705,37	-3,98	7,48E-05	1
201291_s_at	TOP2A	topoisomerase (DNA) II alpha 170kDa	AU159942	14,23	352,24	24,75	5,81E-08	0,00317406
201292_at	TOP2A	topoisomerase (DNA) II alpha 170kDa	AL561834	21,85	389,9	17,84	1,31E-07	0,007159215
201464_x_at	JUN	jun oncogene	BG491844	5669,22	2666,48	-2,13	0,000114	1
201466_s_at	JUN	jun oncogene	NM_002228	3746,62	1114,27	-3,36	5,87E-05	1
201890_at	RRM2	ribonucleotide reductase M2 polypeptide	BE966236	29,48	592,75	20,11	7,88E-07	0,043002588
201904_s_at	CTDSPL	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A)	BF031714	114,84	240,48	2,09	1,81E-05	0,987880928

202112_at	VWF	small phosphatase-like	NM_000552	1852,57	644,3	-2,88	0,000137	1
202340_x_at	NR4A1	von Willebrand factor	NM_002135	1710,19	250,01	-6,84	3,77E-06	0,206082453
202350_s_at	MATN2	nuclear receptor subfamily 4, group A, member 1	NM_002380	6692,29	2405,8	-2,78	0,000103	1
202705_at	CCNB2	matrilin 2	NM_004701	21,53	293,03	13,61	0,000011	0,602042246
202747_s_at	ITM2A	cyclin B2	NM_004867	600,6	292,03	-2,06	0,000114	1
202768_at	FOSB	integral membrane protein 2A	NM_006732	4952,58	584,74	-8,47	0,000005	0,272887993
202871_at	TRAF4	FBJ murine osteosarcoma viral oncogene homolog B	NM_004295	19,93	149,95	7,52	5,87E-05	1
202954_at	UBE2C	TNF receptor-associated factor 4	NM_007019	12,8	202,18	15,79	3,91E-05	1
203029_s_at	PTPRN2	ubiquitin-conjugating enzyme E2C	NM_002847	269,21	123,19	-2,19	0,000027	1
203228_at	PAFAH1B3	protein tyrosine phosphatase, receptor type, N polypeptide 2	NM_002573	87,09	246,61	2,83	1,35E-06	0,073680451
203358_s_at	EZH2	platelet-activating factor acetylhydrolase, isoform 1b, gamma subunit 29kDa	NM_004456	13,19	158,31	12	1,49E-05	0,811473964
203367_at	DUSP14	enhancer of zeste homolog 2 (Drosophila)	NM_007026	801,99	332,13	-2,41	6,04E-05	1
203752_s_at	JUND	dual specificity phosphatase 14	NM_005354	8044,45	4599,98	-1,75	0,000123	1
203755_at	BUB1B	jun D proto-oncogene	NM_001211	14,18	125,41	8,85	3,92E-05	1
203766_s_at	LMOD1	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	NM_012134	314,41	62,5	-5,03	1,69E-05	0,924727338
203980_at	FABP4	leiomodlin 1 (smooth muscle)	NM_001442	2420,85	448,25	-5,4	5,29E-05	1
204170_s_at	CKS2	fatty acid binding protein 4, adipocyte	NM_001827	183,13	699,78	3,82	0,000068	1
204274_at	EBAG9	CDC28 protein kinase regulatory subunit 2	AA812215	2925,89	1820,88	-1,61	3,69E-05	1
204368_at	SLCO2A1	estrogen receptor binding site associated, antigen, 9	NM_005630	996,35	424,22	-2,35	5,02E-05	1
204641_at	NEK2	solute carrier organic anion transporter family, member 2A1	NM_002497	8,13	125,72	15,46	4,52E-05	1
204825_at	MELK	NIMA (never in mitosis gene a)-related kinase 2	NM_014791	8,47	139,43	16,45	2,93E-05	1
205249_at	EGR2	maternal embryonic leucine zipper kinase early growth response 2	NM_000399	1799,53	217,77	-8,26	5,53E-07	0,030171565
205342_s_at	SULT1C2	(Krox-20 homolog, Drosophila)	AF026303	508,16	166,37	-3,05	9,08E-05	1
205357_s_at	AGTR1	sulfotransferase family, cytosolic, 1C, member 2	NM_000685	201,6	29,72	-6,78	0,000123	1
		angiotensin II receptor, type 1						

205413_at	MPPED2	metallophosphoesterase domain containing 2	NM_001584	1392,65	383,88	-3,63	9,58E-06	0,523155106
205449_at	SAC3D1	SAC3 domain containing 1	NM_013299	127,96	362,31	2,83	9,92E-05	1
205554_s_at	DNASE1L3	deoxyribonuclease I-like 3	NM_004944	287,07	40,07	-7,16	8,92E-07	0,048689669
206208_at	CA4	carbonic anhydrase IV	NM_000717	325,1	53,39	-6,09	4,58E-08	0,002499908
206209_s_at	CA4	carbonic anhydrase IV	NM_000717	433,27	59,25	-7,31	1,34E-07	0,00729814
206517_at	CDH16	cadherin 16, KSP-cadherin	NM_004062	276,24	63,33	-4,36	8,46E-05	1
206529_x_at	SLC26A4	solute carrier family 26, member 4	NM_000441	8820,62	3052,43	-2,89	7,53E-05	1
207165_at	HMMR	hyaluronan-mediated motility receptor (RHAMM)	NM_012485	13,54	127,46	9,42	6,89E-06	0,376237336
207828_s_at	CENPF	centromere protein F, 350/400ka (mitosin)	NM_005196	8,74	154,23	17,65	2,77E-05	1
207980_s_at	CITED2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	NM_006079	3503,67	980,87	-3,57	1,01E-05	0,552642819
208002_s_at	ACOT7	acyl-CoA thioesterase 7	NM_007274	20,32	134,39	6,61	0,000119	1
208671_at	SERINC1	serine incorporator 1	AF164794	5662,81	3647,09	-1,55	8,61E-05	1
209101_at	CTGF	connective tissue growth factor	M92934	8371,46	1470,6	-5,69	4,8E-06	0,261968455
209311_at	BCL2L2	BCL2-like 2	D87461	517,77	323,18	-1,6	9,37E-05	1
209357_at	CITED2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	AF109161	5503,58	1763,86	-3,12	2,98E-06	0,162737725
209481_at	SNRK	SNF related kinase	AF226044	1495,39	913,54	-1,64	0,000121	1
209754_s_at	TMPO	thymopoietin	AF113682	32,31	137,9	4,27	0,000119	1
209773_s_at	RRM2	ribonucleotide reductase M2 polypeptide	BC001886	9,79	310,28	31,69	1,44E-07	0,007884026
209814_at	ZNF330	zinc finger protein 330	BC004421	1188,18	764,86	-1,55	0,000104	1
209959_at	NR4A3	nuclear receptor subfamily 4, group A, member 3	U12767	816,33	33,81	-24,14	5,26E-05	1
210052_s_at	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	AF098158	7,48	172,21	23,02	4,87E-05	1
210078_s_at	KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	L39833	2566,59	733,73	-3,5	0,000115	1
211143_x_at	NR4A1	nuclear receptor subfamily 4, group A, member 1	D49728	276,19	76,66	-3,6	4,73E-05	1
211470_s_at	SULT1C2	sulfotransferase family, cytosolic, 1C, member 2	AF186255	1242,47	439,77	-2,83	0,000124	1
212245_at	MCFD2	multiple coagulation factor deficiency 2	AL567779	3492,73	2184,81	-1,6	3,37E-05	1
213618_at	CENTD1	centaurin, delta 1	AB011152	765,87	363,22	-2,11	2,42E-06	0,131907075

213637_at	locus	transcribed locus	BE503392	105,11	208,85	1,99	0,000102	1
213702_x_at	ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	AI934569	10692,84	6072,06	-1,76	0,000113	1
214180_at	MAN1C1	mannosidase, alpha, class 1C, member 1	AW340588	336,06	134,47	-2,5	7,09E-05	1
214307_at	HGD	homogentisate 1,2-dioxygenase (homogentisate oxidase)	AI478172	202,32	52,12	-3,88	3,87E-05	1
214308_s_at	HGD	homogentisate 1,2-dioxygenase (homogentisate oxidase)	AI478172	578,73	178,7	-3,24	4,85E-05	1
214501_s_at	H2AFY	H2A histone family, member Y	AF044286	695,42	1115,73	1,6	1,09E-05	0,592252038
214696_at	MGC14376	hypothetical protein MGC14376 metallophosphoesterase domain containing 2	AF070569	1325,33	450,49	-2,94	2,37E-05	1
215692_s_at	MPPED2	isovaleryl Coenzyme A dehydrogenase	BE645386	277,53	79,96	-3,47	1,91E-05	1
216958_s_at	IVD	protein regulator of cytokinesis 1	AK022777	830,52	410,76	-2,02	0,000085	1
218009_s_at	PRC1	nucleolar and spindle associated protein 1	NM_003981	24,68	239,52	9,7	0,000101	1
218039_at	NUSAP1	kinesin family member 4A	NM_016359	37,36	385,87	10,33	1,73E-05	0,944709766
218355_at	KIF4A	heme binding protein 1	NM_012310	6,5	116,63	17,94	7,06E-05	1
218450_at	HEBP1	Bardet-Biedl syndrome 1	NM_015987	4275,55	2331,65	-1,83	6,54E-05	1
218471_s_at	BBS1	centrosomal protein 55kDa	NM_024649	738,89	413,43	-1,79	8,97E-05	1
218542_at	CEP55	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	NM_018131	7,03	113,37	16,12	7,38E-05	1
218559_s_at	MAFB	serum deprivation response (phosphatidylserine binding protein)	NM_005461	1867,11	726,55	-2,57	2,59E-06	0,141133968
218711_s_at	SDPR	chromosome 13 open reading frame 15	NM_004657	611,84	79,71	-7,68	2,88E-05	1
218723_s_at	C13orf15	mannosidase, alpha, class 1C, member 1	NM_014059	2653,34	1305,13	-2,03	4,05E-05	1
218918_at	MAN1C1	inter-alpha (globulin) inhibitor H5	NM_020379	892,1	307,2	-2,9	2,62E-05	1
219064_at	ITIH5	PDZ binding kinase	NM_030569	221,89	64,23	-3,45	2,85E-06	0,155774509
219148_at	PBK	zinc finger, matrin type 4	NM_018492	6,58	237,91	36,16	6,92E-05	1
219877_at	ZMAT4	asp (abnormal spindle) homolog, microcephaly associated (Drosophila)	NM_024645	197,81	42,89	-4,61	3,75E-05	1
219918_s_at	ASPM	apolipoprotein L domain containing 1	NM_018123	6,65	173,09	26,03	5,62E-05	1
221031_s_at	APOLD1	fumarylacetoacetate hydrolase domain containing 2A	NM_030817	1629,52	658,37	-2,48	1,58E-05	0,860955494
222056_s_at	FAHD2A		AA723370	1325,28	800,03	-1,66	9,85E-05	1

222077_s_at	RACGAP1	Rac GTPase activating protein 1	AU153848	79,45	351,41	4,42	0,000119	1
222581_at	XPR1	xenotropic and polytropic retrovirus receptor	AF089744	361,97	1210,68	3,34	1,82E-06	0,099232458
222608_s_at	ANLN	anillin, actin binding protein	AK023208	7,71	198,13	25,69	7,15E-06	0,390424111
222619_at	ZNF281	zinc finger protein 281	AU150752	191,88	415,59	2,17	7,98E-05	1
222670_s_at	MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	AW135013	1359,71	678,73	-2	7,15E-05	1
222717_at	SDPR	serum deprivation response	BF982174	2058,07	400,75	-5,14	4,52E-06	0,246835129
222833_at	AYTL1	(phosphatidylserine binding protein) acyltransferase like 1	AU154202	400,52	199,24	-2,01	0,000106	1
223960_s_at	C16orf5	chromosome 16 open reading frame 5	AF195661	366,36	187,07	-1,96	4,56E-05	1
224480_s_at	MAG1	lung cancer metastasis-associated protein	BC006236	1161,37	433,07	-2,68	7,76E-05	1
224578_at	RCC2	regulator of chromosome condensation 2	AB040903	305,44	667,89	2,19	0,000019	1
224753_at	CDCA5	cell division cycle associated 5	BE614410	13,46	164,36	12,21	0,000076	1
224822_at	DLC1	deleted in liver cancer 1	AA524250	536,43	263,65	-2,03	9,08E-05	1
224892_at	PLDN	pallidin homolog (mouse)	BF680495	2518,18	1613,03	-1,56	4,06E-05	1
225311_at	IVD	isovaleryl Coenzyme A dehydrogenase	AA081349	832,08	416,08	-2	3,32E-05	1
225522_at	AAK1	AP2 associated kinase 1	AW628987	2859,19	1445,69	-1,98	1,15E-05	0,629553019
225589_at	SH3RF1	SH3 domain containing ring finger 1	AB040927	851,8	451,68	-1,89	1,72E-05	0,941235327
225677_at	BCAP29	B-cell receptor-associated protein 29	AW152589	2046,4	1212,31	-1,69	0,000117	1
226120_at	TTC8	tetratricopeptide repeat domain 8	AW293939	875,98	443,22	-1,98	5,42E-05	1
226914_at	ARPC5L	actin related protein 2/3 complex, subunit 5-like	AU158936	201,85	352,59	1,75	0,000023	1
226915_s_at	ARPC5L	actin related protein 2/3 complex, subunit 5-like	AU158936	464,86	723,91	1,56	2,56E-05	1
228368_at	ARHGAP20	Rho GTPase activating protein 20	A1936560	217,57	62,41	-3,49	1,88E-06	0,102523365
229490_s_at	Transcribed locus	transcribed locus	AW271106	20,83	199,05	9,55	0,000106	1
231882_at	FLJ39632	hypothetical LOC642477	AL530703	29,9	177,09	5,92	3,95E-05	1
235228_at	CCDC85A	coiled-coil domain containing 85A	A1376433	286,22	118,1	-2,42	1,28E-05	0,698180214
235746_s_at	PLA2R1	phospholipase A2 receptor 1, 180kDa	BE048919	399,46	99,47	-4,02	1,3E-06	0,071072621
242881_x_at	Transcribed locus	Clone HLS_IMAGE_626842 mRNA sequence	BG285837	15,17	310,67	20,47	2,86E-05	1

Table 20 (FEA_vs_FA_240) lists the 240 probe sets found to be differentially expressed in the class comparison between FA and FEA. Gene Ontology annotation is shown as is the result of several multiple testing correction algorithms. The base line is set to the FA class and the experiment is set to the FEA class. Raw p is the p-value of Student's t-test. Genes are filtered to have a Benjamini Hochberg corrected p values below 0.05, an absolute fold change above 1.5 as well as an absolute change in expression levels of more than 100.

Probe set	Symbol and gene	Accession	Baseline mean	Experiment mean	Fold change	Raw p	Hochberg
1552256_a_at	SCARB1: scavenger receptor class B, member 1	NM_005505	317,24	671,11	2,12	0,000362	1
1553148_a_at	SNX13: sorting nexin 13	R75838	112,8	272,65	2,42	0,000195	1
1554128_at	ADIG: adipogenin	BC029594	29,4	130,1	4,42	4,07E-05	1
1555789_s_at	NUP98 /// PHF23: nucleoporin 98kDa /// PHD finger protein 23	AY099328	146,26	291,68	1,99	0,000177	1
1556180_at	LOC729678: hypothetical protein LOC729678	BE646146	286,41	604,5	2,11	0,000121	1
1558953_s_at	CEP164: centrosomal protein 164kDa	BC000602	435,3	225,71	-1,93	0,000164	1
1559881_s_at	ZNF12: zinc finger protein 12	BM463827	276,65	544,57	1,97	8,07E-05	1
1561079_at	ANKRD28: ankyrin repeat domain 28	BC035170	154,79	49,64	-3,12	0,000157	1
200011_s_at	ARF3: ADP-ribosylation factor 3	NM_001659	3089,51	4825,6	1,56	1,01E-05	5
200617_at	KIAA0152: KIAA0152	NM_014730	2397,81	3894	1,62	9,07E-05	1
200694_s_at	DDX24: DEAD (Asp-Glu-Ala-Asp) box polypeptide 24	NM_020414	1833,81	3485,62	1,9	1,86E-06	8
200734_s_at	ARF3: ADP-ribosylation factor 3	BG341906	721,7	1193,92	1,65	0,000262	1
200891_s_at	SSR1: signal sequence receptor, alpha (translocon-associated protein alpha)	NM_003144	6074,55	3978,15	-1,53	0,000156	1
200894_s_at	FKBP4: FK506 binding protein 4, 59kDa	AA894574	308,75	696,56	2,26	4,14E-06	9
200895_s_at	FKBP4: FK506 binding protein 4, 59kDa	NM_002014	1255,38	2240,59	1,78	0,000105	1
200907_s_at	PALLD: palladin, cytoskeletal associated protein	AU157932	2112,62	1198,55	-1,76	0,000354	1
200948_at	MLF2: myeloid leukemia factor 2	NM_005439	1465,44	2469,78	1,69	3,64E-05	1
200950_at	ARPC1A: actin related protein 2/3 complex, subunit 1A, 41kDa	NM_006409	1470,41	2740,01	1,86	1,99E-05	1
201067_at	PSMC2: proteasome (prosome, macropain) 26S subunit, 621	BF215487	621	1011,78	1,63	0,000164	1

ATPase, 2									
201068_s_at	PSMC2: proteasome (prosome, macropain) 26S subunit,	NM_002803	2745,47	4182,09	1,52	0,000256	1		
201191_at	ATPase, 2	H15647	586,4	901,89	1,54	0,00023	1		
201276_at	PITPNA: phosphatidylinositol transfer protein, alpha	AF267863	664,74	1032,05	1,55	0,000176	1		
	RAB5B: RAB5B, member RAS oncogene family						0,0656802		
201285_at	MKRN1: makorin, ring finger protein, 1	NM_013446	1084,73	1720,09	1,59	1,2E-06	1		
201331_s_at	STAT6: signal transducer and activator of transcription 6, interleukin-4 induced	BC004973	555,53	885,87	1,59	7,2E-05	1		
201405_s_at	COPS6: COP9 constitutive photomorphogenic homolog subunit 6 (Arabidopsis)	NM_006833	975,3	1546,7	1,59	8,63E-05	1		
201440_at	DDX23: DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	NM_004818	285,47	429,39	1,5	5,72E-06	1	0,3124246	
201526_at	ARF5: ADP-ribosylation factor 5	NM_001662	316,44	557,74	1,76	4,17E-05	1		
201541_s_at	ZNHIT1: zinc finger, HIT type 1	NM_006349	975,99	1592,94	1,63	1,69E-05	2	0,9224808	
201861_s_at	LRRFIP1: leucine rich repeat (in FLII) interacting protein 1	BF965566	405,55	234,05	-1,73	5,25E-06	0,2866033	0,9141147	
201962_s_at	RNF41: ring finger protein 41	NM_005785	128,19	255,95	2	1,68E-05	6		
202031_s_at	WIP2: WD repeat domain, phosphoinositide interacting 2	NM_015610	862,89	1438,07	1,67	2,42E-05	1		
202185_at	PLOD3: procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	NM_001084	675,13	1172,2	1,74	0,000102	1		
202246_s_at	CDK4: cyclin-dependent kinase 4	NM_000075	1183,87	1898,11	1,6	0,000312	1		
202290_at	PDAP1: PDGFA associated protein 1	NM_014891	551,66	873,96	1,58	3,24E-05	1		
202365_at	UNC119B: unc-119 homolog B (C. elegans)	BC004815	1143,06	1793,9	1,57	0,000187	1		
202487_s_at	H2AFV: H2A histone family, member V	NM_012412	909,26	1548,22	1,7	2,16E-05	1		
202518_at	BCL7B: B-cell CLL/lymphoma 7B	NM_001707	302,92	486,78	1,61	0,000307	1		
202556_s_at	MCRS1: microspherule protein 1	NM_006337	351,61	579,39	1,65	0,000132	1		
202592_at	BLOC1S1: biogenesis of lysosome-related organelles complex-1, subunit 1	NM_001487	1318,09	2096,44	1,59	0,000308	1		
							0,0607999		
202858_at	U2AF1: U2 small nuclear RNA auxiliary factor 1	NM_006758	2920,01	4944,59	1,69	1,11E-06	6		
203415_at	PDCCD6: programmed cell death 6	NM_013232	1433,82	2713,15	1,89	6,44E-05	1		
203454_s_at	ATOX1: ATX1 antioxidant protein 1 homolog (yeast)	NM_004045	754,43	1488,43	1,97	4,58E-05	1		
203502_at	BPGM: 2,3-bisphosphoglycerate mutase	NM_001724	168,43	356,67	2,12	0,000101	1		
203533_s_at	CUL5: cullin 5	NM_003478	906,24	572,23	-1,58	0,000278	1		

203629_s_at	COG5: component of oligomeric golgi complex 5	AU152134	500,18	810,61	1,62	0,000256	1
203630_s_at	COG5: component of oligomeric golgi complex 5	NM_006348	568,17	994,55	1,75	0,000254	1
203733_at	DEXI: dexamethasone-induced transcript	NM_014015	350,43	715,8	2,04	7,15E-07	0,0390189
203812_at	CDNA clone IMAGE:5922621	AB011538	169,21	329,75	1,95	0,000335	6
204031_s_at	PCBP2: poly(rC) binding protein 2	NM_005016	5192,07	8386,48	1,62	0,000365	1
204067_at	SUOX: sulfite oxidase	AA129776	377,78	694,2	1,84	1,02E-05	0,5565126
204202_at	IQCE: IQ motif containing E	NM_017604	272,28	506,29	1,86	8,2E-05	9
204488_at	DOLK: dolichol kinase	NM_014908	138,56	265,75	1,92	0,000306	1
204796_at	EML1: echinoderm microtubule associated protein like 1	A1825937	96,65	207,03	2,14	1,95E-05	1
204797_s_at	EML1: echinoderm microtubule associated protein like 1	NM_004434	187,33	358,09	1,91	1,12E-05	0,6089439
205194_at	PSPH: phosphoserine phosphatase	NM_004577	216,15	596,65	2,76	9,52E-07	9
206087_x_at	HFE: hemochromatosis	NM_000410	206,77	352,93	1,71	0,000212	0,0519848
206355_at	GNAL: guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type	R20102	104,73	413,35	3,95	0,000428	8
206845_s_at	RNF40: ring finger protein 40	NM_014771	200,21	335,78	1,68	1,88E-05	1
207614_s_at	CUL1: cullin 1	NM_003592	611,31	1025,09	1,68	1,79E-05	0,9787673
207628_s_at	WBSR22: Williams Beuren syndrome chromosome region 22	NM_017528	938,66	1575,88	1,68	0,000154	6
207805_s_at	PSMD9: proteasome (prosome, macropain) 26S subunit, non-ATPase, 9	NM_002813	733,8	1124,92	1,53	3,44E-05	1
208146_s_at	CPVL: carboxypeptidase, vitellogenic-like	NM_031311	1365,53	764,06	-1,79	0,000234	1
208445_s_at	BAZ1B: bromodomain adjacent to zinc finger domain, 1B	NM_023005	512,21	817,2	1,6	0,000451	1
208447_s_at	PRPS1: phosphoribosyl pyrophosphate synthetase 1	NM_002764	188,18	394,77	2,1	7,66E-05	1
208503_s_at	GATAD1: GATA zinc finger domain containing 1	NM_021167	876,52	1613,32	1,84	0,00026	1
208722_s_at	ANAPC5: anaphase promoting complex subunit 5	BC001081	619,09	942,26	1,52	9,53E-06	0,5200459
208792_s_at	CLU: clusterin	M25915	9702,93	3894,52	-2,49	0,000388	9
208828_at	POLE3: polymerase (DNA directed), epsilon 3 (p17 subunit)	BC004170	320,99	530,57	1,65	0,000121	1
208912_s_at	CNP: 2',3'-cyclic nucleotide 3' phosphodiesterase	BC001362	463,26	731,77	1,58	3,5E-05	1

208928_at	POR: P450 (cytochrome) oxidoreductase	AF258341	1218,17	3060,85	2,51	8,19E-05	1
209015_s_at	DNAJB6: DnaJ (Hsp40) homolog, subfamily B, member 6	BC002446	157,12	325,85	2,07	9,34E-05	1
209150_s_at	TM9SF1: transmembrane 9 superfamily member 1	U94831	950,34	1462,8	1,54	0,000251	1
209247_s_at	ABCF2: ATP-binding cassette, sub-family F (GCN20), member 2	BC001661	211,05	470,81	2,23	8,16E-06	0,4452717
209256_s_at	KIAA0265: KIAA0265 protein	AF277177	411,74	737,4	1,79	7,23E-05	1
209318_x_at	PLAGL1: pleiomorphic adenoma gene-like 1	BG547855	165,26	46,32	-3,57	0,000137	1
209440_at	PRPS1: phosphoribosyl pyrophosphate synthetase 1	BC001605	791,59	1526,85	1,93	1,77E-05	0,9657768
209482_at	POP7: processing of precursor 7, ribonuclease P/MRP subunit (S. cerevisiae)	BC001430	229,51	374,73	1,63	7,98E-05	1
209515_s_at	RAB27A: RAB27A, member RAS oncogene family	U38654	1153,2	416,82	-2,77	0,000241	1
209531_at	GSTZ1: glutathione transferase zeta 1 (maleylacetoacetate isomerase)	BC001453	286,35	491,89	1,72	0,000102	1
209623_at	MCCC2: methylcrotonoyl-Coenzyme A carboxylase 2 (beta)	AW439494	1185,5	2181,99	1,84	0,000112	1
209681_at	SLC19A2: solute carrier family 19 (thiamine transporter), member 2	AF153330	520,06	242,06	-2,15	0,000446	1
209796_s_at	TMEM4: transmembrane protein 4	BC001027	1397,95	2390,86	1,71	3,35E-06	0,1830495
209898_x_at	ITSN2: intersectin 2	U61167	2938,52	1778,04	-1,65	0,000242	9
209917_s_at	TP53AP1: TP53 activated protein 1	BC002709	301,5	802,4	2,66	1,16E-05	1
210241_s_at	TP53AP1: TP53 activated protein 1	AB007458	396,33	821,32	2,07	0,000305	5
210529_s_at	FAM115A /// FAM115B: family with sequence similarity 115, member A /// family with sequence similarity 115, member B	BC000609	139,76	333,86	2,39	8,48E-05	1
210534_s_at	EPPB9: B9 protein	BC002944	356,17	621,69	1,75	0,000404	1
210788_s_at	DHRS7: dehydrogenase/reductase (SDR family) member 7	AF126782	2269,8	3694,16	1,63	8,82E-05	1
210886_x_at	TP53AP1: TP53 activated protein 1	AB007457	445,87	944,02	2,12	0,000172	1
212010_s_at	CDV3: CDV3 homolog (mouse)	AK025647	5012,75	3128,31	-1,6	0,000374	1
212056_at	KIAA0182: KIAA0182	D80004	652,09	1067,53	1,64	0,000228	1
212114_at	LOC552889: hypothetical LOC552889	BE967207	1193,38	2087,23	1,75	1,71E-05	0,9317881
212205_at	H2AFV: H2A histone family, member V	AA534860	2233,7	3364,03	1,51	1,24E-05	2
							0,6782780
							9

212206_s_at	H2AFV: H2A histone family, member V	BF343852	555,93	1068,68	1,92	3,24E-07	0,0176682
212403_at	UBE3B: ubiquitin protein ligase E3B	AI749193	371,58	559,44	1,51	3,55E-05	5
212685_s_at	TBL2: transducin (beta)-like 2	AI608789	1350,61	2109,65	1,56	0,000258	1
212708_at	MSL-1: male-specific lethal-1 homolog	AV721987	834,96	1515,31	1,81	3,32E-06	0,1812935
212739_s_at	NME4: non-metastatic cells 4, protein expressed in	AL523860	488,66	949,94	1,94	0,000168	9
212746_s_at	CEP170: centrosomal protein 170kDa	AA126789	335,02	190,51	-1,76	7,35E-05	1
212782_x_at	POLR2J: polymerase (RNA) II (DNA directed) polypeptide J, 13.3kDa	BG335629	2413,22	4041,68	1,67	5,06E-05	1
212814_at	KIAA0828: adenosylhomocysteinase 3	AB020635	671,9	1341,67	2	5,26E-05	1
212825_at	PAXIP1: PAX interacting (with transcription-activation domain) protein 1	AI357401	281,37	472,02	1,68	0,00014	1
213234_at	KIAA1467: KIAA1467	AB040900	134,73	324,06	2,41	0,000248	1
213275_x_at	CTSB: cathepsin B	W47179	14655,89	8062,07	-1,82	0,000168	1
213504_at	COPS6: COP9 constitutive photomorphogenic homolog subunit 6 (Arabidopsis)	W63732	351,49	661,21	1,88	1,22E-05	0,6660875
213508_at	C14orf147: chromosome 14 open reading frame 147	AA142942	530,11	981,64	1,85	4,39E-05	2
213861_s_at	FAM119B: family with sequence similarity 119, member B	N67741	142,52	308,34	2,16	1,01E-05	0,5502469
214151_s_at	CCPG1: cell cycle progression 1	AU144243	539,35	299,82	-1,8	0,00018	9
214152_at	CCPG1: cell cycle progression 1	AU144243	583,35	315,89	-1,85	0,000136	1
214746_s_at	ZNF467: zinc finger protein 467	BE549732	100,17	237,75	2,37	1,38E-05	0,7501547
214794_at	PA2G4: proliferation-associated 2G4, 38kDa	BF669264	402,17	704,39	1,75	0,000416	7
215145_s_at	CNTNAP2: contactin associated protein-like 2 POLR2J2 /// POLR2J3 /// POLR2J4: polymerase (RNA) II (DNA directed) polypeptide J, 13.3kDa pseudogene /// DNA directed RNA polymerase II polypeptide J-related /// RPB11b2 protein	AC005378	29,61	135,23	4,57	0,000448	1
216242_x_at	CHCHD2: coiled-coil-helix-coiled-coil-helix domain containing 2	AW402635	362,64	587,16	1,62	0,000272	1
217720_at		NM_016139	5741,76	8823,76	1,54	4,46E-05	1
217726_at	COPZ1: coatamer protein complex, subunit zeta 1	NM_016057	1573,49	2434,43	1,55	7,79E-06	0,4252822
217990_at	GMPR2: guanosine monophosphate reductase 2	NM_016576	803,13	1244,98	1,55	0,000253	7
							1

218034_at	FIS1: fission 1 (mitochondrial outer membrane) homolog (S. cerevisiae)	NM_016068	2465,4	4265,74	1,73	2,97E-07	0,0162176
218045_x_at	PTMS: parathymosin	NM_002824	283,74	517,97	1,83	0,000335	6
218220_at	C12orf10: chromosome 12 open reading frame 10	NM_021640	333,04	502,45	1,51	0,000354	1
218321_x_at	STYXL1: serine/threonine/tyrosine interacting-like 1	NM_016086	252,71	451,82	1,79	0,000339	1
218328_at	COQ4: coenzyme Q4 homolog (S. cerevisiae)	NM_016035	610,81	1405,38	2,3	7,33E-05	1
218347_at	TYW1: tRNA-yW synthesizing protein 1 homolog (S. cerevisiae)	NM_018264	392,33	681,84	1,74	0,000364	1
218417_s_at	FLJ20489: hypothetical protein FLJ20489	NM_017842	406,95	699,35	1,72	0,000209	1
218460_at	HEATR2: HEAT repeat containing 2	NM_017802	416,56	645,04	1,55	0,000282	1
218601_at	URG4: up-regulated gene 4	NM_017920	266,86	467,79	1,75	8,13E-05	1
218654_s_at	MRPS33: mitochondrial ribosomal protein S33	NM_016071	1792,86	3677,49	2,05	0,000174	1
218667_at	PJA1: praja 1	NM_022368	538,28	985,72	1,83	0,000461	1
218984_at	PUS7: pseudouridylate synthase 7 homolog (S. cerevisiae)	NM_019042	338,87	815,43	2,41	2,73E-05	1
219043_s_at	LOC285359 /// LOC644850 /// PDCL3: phosducin-like 3	NM_024065	765,11	508,52	-1,5	7,62E-05	1
219569_s_at	phosducin-like 3 pseudogene /// similar to phosducin-like 3	NM_025246	126,22	256,45	2,03	0,000452	1
219682_s_at	TMEM22: transmembrane protein 22	NM_016569	1674,88	3231,87	1,93	2,5E-05	1
219798_s_at	TBX3: T-box 3 (ulnar mammary syndrome)	NM_019606	911,22	1702,38	1,87	5,98E-06	0,3264234
	MEPCE: methylphosphate capping enzyme						0,4031357
220261_s_at	ZDHHC4: zinc finger, DHC-type containing 4	NM_018106	480,91	1048,83	2,18	7,39E-06	8
220617_s_at	ZNF532: zinc finger protein 532	NM_018181	408,34	233,6	-1,75	0,000285	1
220690_s_at	DHRS7B: dehydrogenase/reductase (SDR family) member 7B	NM_015510	356,18	562,86	1,58	0,000386	1
220949_s_at	C7orf49: chromosome 7 open reading frame 49	NM_024033	191,37	316,8	1,66	0,000272	1
221247_s_at	WBSCR16: Williams-Beuren syndrome chromosome region 16	NM_030798	229,47	373,78	1,63	7,57E-05	1
221447_s_at	GLT8D2: glycosyltransferase 8 domain containing 2	NM_031302	284,46	1492,45	5,25	2,16E-05	1
221580_s_at	JOSD3: Josephin domain containing 3	BC001972	724,86	434,19	-1,67	9,21E-05	1
221596_s_at	DKFZP564O0523: hypothetical protein DKFZp564O0523	AL136619	212,04	424,02	2	3,44E-05	1
221654_s_at	USP3: ubiquitin specific peptidase 3	AF077040	395,53	235,47	-1,68	0,00039	1
221686_s_at	RECQL5: RecQ protein-like 5	AL136869	259,99	488,42	1,88	0,000124	1
221893_s_at	ADCK2: aarF domain containing kinase 2	N32831	197,85	375,48	1,9	0,000231	1
222294_s_at	CDNA clone IMAGE:5745639	AW971415	358,16	110,79	-3,23	7,93E-05	1

222512_at	NUB1: negative regulator of ubiquitin-like proteins 1	AF300717	530,77	850,44	1,6	0,000318	1
222742_s_at	RABL5: RAB, member RAS oncogene family-like 5	AW026449	749,47	1619,64	2,16	1,92E-05	1
222778_s_at	WHSC1: Wolf-Hirschhorn syndrome candidate 1	AW024870	59,7	179,79	3,01	5E-05	1
222798_at	PTER: phosphotriesterase related	BF112019	580,28	303,4	-1,91	0,000456	1
222867_s_at	MED31: mediator complex subunit 31	AV760596	436,75	688,36	1,58	6,31E-05	1
222995_s_at	RHBDD2: rhomboid domain containing 2	AF226732	1373,85	3033,65	2,21	1,97E-05	1
223007_s_at	C9orf5: chromosome 9 open reading frame 5	AA495988	1580,79	2477,73	1,57	0,000395	1
223065_s_at	STARD3NL: STARD3 N-terminal like	BC003074	1000,56	1659,25	1,66	0,00015	1
223081_at	PHF23: PHD finger protein 23	BC002509	549,37	967,08	1,76	6,37E-05	1
223109_at	TRUB2: TruB pseudouridine (psi) synthase homolog 2 (E. coli)	BC001457	300,67	489,95	1,63	0,000429	1
223139_s_at	DHX36: DEAH (Asp-Glu-Ala-His) box polypeptide 36	BE501133	2062,95	1359,82	-1,52	0,000273	1
223148_at	PLGS: phosphatidylinositol glycan anchor biosynthesis, class S	BC001319	116,74	274,17	2,35	9,45E-06	9
223162_s_at	KIAA1147: KIAA1147	AF116707	593,31	1027,42	1,73	7,51E-05	1
223367_at	WBSR18: Williams Beuren syndrome chromosome region 18	BC005056	441,68	807,39	1,83	1,33E-05	3
223420_at	DNAJC14: DnaJ (Hsp40) homolog, subfamily C, member 14	AA156470	239,12	396,59	1,66	0,000317	1
223433_at	C7orf36: chromosome 7 open reading frame 36	AF226046	255,4	398,02	1,56	1,15E-05	4
223457_at	COPG2: coatomer protein complex, subunit gamma 2	AB047847	80,48	197,63	2,46	9,58E-05	1
223459_s_at	C1orf56: chromosome 1 open reading frame 56	BE222214	187,46	422,37	2,25	6,9E-06	6
224415_s_at	HINT2: histidine triad nucleotide binding protein 2	AF356515	897,22	1516,07	1,69	5,88E-05	1
224558_s_at	MALAT1: metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)	AI446756	3531,45	1134,97	-3,11	9,26E-05	1
224626_at	SLC35A4: solute carrier family 35, member A4	BE618656	501,3	819,89	1,64	7,49E-05	1
224682_at	ANKIB1: ankyrin repeat and IBR domain containing 1	AA253488	519,79	867,5	1,67	0,000399	1
224688_at	C7orf42: chromosome 7 open reading frame 42	BE962299	794,29	1294,79	1,63	4,99E-06	1
224732_at	CTF8: chromosome transmission fidelity factor 8 homolog (S. cerevisiae)	AI309784	694,81	1162,45	1,67	0,000249	1
224748_at	WDR68: WD repeat domain 68	AK025925	934,54	1469,09	1,57	0,000111	1
224767_at	RPL37: Ribosomal protein L37	AL044126	211,17	463,77	2,2	0,000259	1

224826_at	RP5-1022P6.2: hypothetical protein KIAA1434	AK001947	618,12	378,55	-1,63	0,000284	1
224879_at	C9orf123: chromosome 9 open reading frame 123 VKORC1L1: vitamin K epoxide reductase complex, subunit 1-like 1	BF315994	1161,53	1785,17	1,54	8,94E-05	1
224881_at		AV724827	204,66	355	1,73	7,14E-05	1
225002_s_at	SUMF2: sulfatase modifying factor 2	BE349022	1518,18	2553	1,68	0,000196	1
225021_at	ZNF532: zinc finger protein 532	AA861416	268	148,09	-1,81	0,000393	1
225134_at	SPRYD3: SPRY domain containing 3	AF131774	528,57	852,31	1,61	0,000109	1
225193_at	KIAA1967: KIAA1967	BC003172	104,71	254,78	2,43	7,38E-05	1
225260_s_at	MRPL32: mitochondrial ribosomal protein L32 TMED8: transmembrane emp24 protein transport domain containing 8	AL551823	1853,92	3042,63	1,64	1,16E-05	0,6354157 7
225343_at		AK025695	1160,31	1910,89	1,65	0,000184	1
225430_at	15E1.2: Hypothetical protein LOC283459	AA541697	538,51	880,86	1,64	6,83E-05	1
225484_at	TSGA14: testis specific, 14	AW157525	89,82	270,17	3,01	2,98E-06	0,1625989 2
225485_at	TSGA14: testis specific, 14	AJ278890	112,88	297,12	2,63	1,29E-05	0,7021679 3
225540_at	MAP2: microtubule-associated protein 2	BF342661	515,12	106,87	-4,82	3E-05	1
225544_at	TBX3: T-box 3 (ulnar mammary syndrome)	AI806338	1777,72	2794,59	1,57	0,000354	1
225772_s_at	C12orf62: chromosome 12 open reading frame 62	BF203664	1138,31	1773,38	1,56	0,000238	1
225804_at	CYB5D2: cytochrome b5 domain containing 2	BE044480	362,65	694,68	1,92	0,000273	1
225837_at	C12orf32: chromosome 12 open reading frame 32	AL577977	171,23	300,89	1,76	0,00045	1
225851_at	FNTB: farnesyltransferase, CAAX box, beta DCP1B: DCP1 decapping enzyme homolog B (S. cerevisiae)	BF131248	170,17	274,07	1,61	0,000409	1
226093_at		AW204088	346,26	554,73	1,6	8,81E-05	1
226146_at	CDNA clone IMAGE:5294560	BE503186	1294,57	2229,96	1,72	8,67E-05	1
226241_s_at	MRPL52: mitochondrial ribosomal protein L52	BG497211	603,32	924,37	1,53	0,00033	1
226434_at	C7orf47: chromosome 7 open reading frame 47	BF000655	191,12	297,79	1,56	0,000414	1
226515_at	CCDC127: coiled-coil domain containing 127	AL577758	249,83	386,53	1,55	0,000364	1
226529_at	TMEM106B: transmembrane protein 106B	BF513060	2670,86	4142,77	1,55	0,00018	1
226546_at	CDNA clone IMAGE:5268696	BG477064	397,27	787,65	1,98	2,47E-05	1
226780_s_at	C7orf55: chromosome 7 open reading frame 55	BF540829	856,79	1436,22	1,68	1,17E-05	0,6396105 4
226781_at	C7orf55: chromosome 7 open reading frame 55	BF540829	587,02	971,58	1,66	2,3E-05	1

226807_at	ZFP1: zinc finger protein 1 homolog (mouse)	AL038511	191,65	322,39	1,68	0,000442	1
226875_at	DOCK11: dedicator of cytokinesis 11	AI742838	664,92	241,35	-2,76	1,52E-05	0,8299518
226938_at	WDR21A: WD repeat domain 21A	AA160604	219,45	373,21	1,7	0,000136	3
226970_at	FBXO33: F-box protein 33	AI690694	337,27	592,86	1,76	0,000195	1
227035_x_at	LOC441212: retinitis pigmentosa 9 pseudogene	BE670798	315,56	538,64	1,71	0,000265	1
227070_at	GLT8D2: glycosyltransferase 8 domain containing 2 IMMP2L: IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)	W63754	656,96	2677,02	4,07	0,000267	1
227153_at		AI784580	415,01	689,79	1,66	6,45E-05	1
227318_at	CDNA FLJ39162 fis, clone OCBBF2002376	AL359605	453,15	1206,27	2,66	0,000296	1
227388_at	TUSC1: tumor suppressor candidate 1	AA479016	575,3	1097,9	1,91	0,000384	1
227391_x_at	LRRFIP1: leucine rich repeat (in FLI) interacting protein 1	BE674143	1923,21	1230,29	-1,56	4,22E-05	1
227572_at	USP30: ubiquitin specific peptidase 30	AA528138	388,62	654,52	1,68	1,58E-05	0,8617127
228087_at	CCDC126: coiled-coil domain containing 126	AK026684	186,65	398,13	2,13	0,00022	6
228899_at	CUL1: Cullin 1 Full-length cDNA clone CS0DC006YB07 of Neuroblastoma	AI870903	65,09	177,34	2,72	0,00035	1
229083_at	Cot 25-normalized of Homo sapiens (human)	AI672356	278,55	495,94	1,78	0,000159	1
229126_at	TMEM19: transmembrane protein 19	AI597651	358,29	636,06	1,78	0,000302	1
229576_s_at	TBX3: T-box 3 (ulnar mammary syndrome)	N29712	263,02	539,02	2,05	2,42E-05	1
230277_at	Hs.115659.1	AI806865	43,45	149,41	3,44	1,74E-05	0,9510310
230669_at	RASA2: RAS p21 protein activator 2	W38444	242,94	140,86	-1,72	0,000332	4
231847_at	ANKRD54: ankyrin repeat domain 54	Z97630	254,18	431,63	1,7	0,000192	1
232053_x_at	RHBDD2: rhomboid domain containing 2 NDUFC1: NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	AL533352	682,57	1522,93	2,23	0,000133	1
232146_at		AK023115	127,22	248,81	1,96	0,000331	1
233982_x_at	STYXL1: serine/threonine/tyrosine interacting-like 1	AF188204	222,25	408,76	1,84	0,000419	1
234135_x_at	CDNA FLJ11590 fis, clone HEMBA1003758	AK021652	197,66	97,33	-2,03	0,000334	1
234302_s_at	ALKBH5: alkB, alkylation repair homolog 5 (E. coli)	AL137263	1156,2	1765,96	1,53	6,44E-05	1
235010_at	LOC729013: hypothetical protein LOC729013	AA833832	337,46	685,24	2,03	6,95E-05	1
235158_at	FLJ14803: hypothetical protein FLJ14803	AI807036	478,62	878,17	1,83	0,00013	1
235282_at	CDNA clone IMAGE:5206119	BF447113	154,56	331,91	2,15	0,000134	1
235371_at	GLT8D4: glycosyltransferase 8 domain containing 4	AI452595	938,06	430,24	-2,18	3,33E-05	1

235587_at	LOC202781: hypothetical protein LOC202781	BG400596	327,68	553,34	1,69	0,000406	1
236404_at	Transcribed locus	AW197320	224,38	63,87	-3,51	0,000279	1
238558_at	Transcribed locus	AI445833	172,82	61,49	-2,81	0,000303	1
238967_at	Transcribed locus	AI924046	284,35	31,38	-9,06	5,48E-05	1
239065_at	Transcribed locus	AA088446	244,59	521,48	2,13	0,00014	1
244414_at	Hs.222120.0	AI148006	232,01	64,62	-3,59	0,000161	1
244587_at	Hs.105820.0	AA534039	135,55	261,06	1,93	2,08E-06	0,1134878
35254_at	TRAFD1: TRAF-type zinc finger domain containing 1	AB007447	110,54	220,47	1,99	1,74E-05	0,9478899
35846_at	THRA: thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)	M24899	169,72	318,84	1,88	4,14E-05	9
40465_at	DDX23: DEAD (Asp-Glu-Ala-Asp) box polypeptide 23	AF026402	315,96	476,89	1,51	9,7E-06	1
44120_at	ADCK2: aarF domain containing kinase 2	AI879381	230,53	353,03	1,53	0,0003	0,5291045
44654_at	G6PC3: glucose 6 phosphatase, catalytic, 3	AI669655	1339,09	2100,58	1,57	3,67E-06	0,2003772
47083_at	C7orf26: chromosome 7 open reading frame 26	AI280108	165,71	283,41	1,71	5,4E-05	3
48106_at	FLJ20489: hypothetical protein FLJ20489	H14241	610,2	1002,1	1,64	0,000199	1
59437_at	C9orf116: chromosome 9 open reading frame 116	AI830563	62,78	182,86	2,91	1,24E-06	0,0677172
							8

Table 21 (FC_vs_FEA_512) lists the 512 probe sets found to be differentially expressed in the class comparison between FEA and FC.

Gene Ontology annotation is shown as is the result of several multiple testing correction algorithms. The base line is set to the FEA class and the experiment is set to the FC class. Raw p is the p-value of Students t-test. Genes are filtered to have a Benjamini Hochberg corrected p values below 0.05, an absolute fold change above 1.5 as well as an absolute change in expression levels of more than 100.

Probe set	Gene and symbol	Accession	Baseline mean	Experiment mean	Fold change	Raw p-value	Hochberg
1552711_a_at	CYB5D1: cytochrome b5 domain containing 1 TIPRL: TIP41, TOR signaling pathway regulator-like (S. cerevisiae)	NM_144607	585,61	293,09	-2	0,000185	1
1553677_a_at		NM_152902	151,62	302,33	1,99	0,000201	1
1554128_at	ADIG: adipogenin	BC029594	130,1	15,5	-8,39	2,79E-07	0,01525813
1554462_a_at	DNAJB9: DnaJ (Hsp40) homolog, subfamily B, member 9	AF115512	1005,41	430,23	-2,34	0,000694	1
1555762_s_at	RBM15: RNA binding motif protein 15	AF364036	176,26	332,24	1,88	0,000307	1
1555910_at	PTCD2: pentatricopeptide repeat domain 2	AK056761	215,81	113,82	-1,9	0,000588	1
1556060_a_at	KIAA1702: KIAA1702 protein	AK027074	304,34	669,98	2,2	0,000412	1
1556151_at	ITFG1: Integrin alpha FG-GAP repeat containing 1	AI077660	1511,34	982,71	-1,54	0,000835	1
1557107_at	LOC286002: hypothetical protein LOC286002	BC037315	3666,75	801,69	-4,57	0,000211	1
1557776_at	CDNA clone IMAGE:4813089	BC030768	166,37	59	-2,82	2,3E-05	1
1557965_at	MTERFD2: MTERF domain containing 2	AL566167	716,37	373,27	-1,92	0,000228	1
1557966_x_at	MTERFD2: MTERF domain containing 2	AL566167	294,56	150,64	-1,96	0,000652	1
1558662_s_at	BANK1: B-cell scaffold protein with ankyrin repeats 1	BG200452	341,33	103,07	-3,31	0,000643	1
1558685_a_at	LOC158960: hypothetical protein BC009467	BC009467	742,87	346,74	-2,14	0,000435	1
1559156_at	MRNA; cDNA DKFZp686B1142 (from clone DKFZp686B1142)	BC036508	197,78	452,62	2,29	0,00072	1
1559391_s_at	Partial mRNA; ID EE2-8E	AI084451	14,97	125,93	8,41	1,16E-05	0,6334047
1560271_at	CDNA clone IMAGE:4797534	BC030757	82,51	253,27	3,07	0,000417	1
1560622_at	CDNA FLJ20196 fis, clone COLF0944	AK000203	124,68	376,21	3,02	0,000714	1
1560926_at	Full length insert cDNA clone YR43G06 LOC646736: similar to Alu subfamily SX sequence contamination warning entry	AF085924	32,06	135,55	4,23	8,57E-05	1
1570289_at		BC017935	399	107,69	-3,7	8,61E-05	1
200011_s_at	ARF3: ADP-ribosylation factor 3	NM_001659	4825,6	2725,45	-1,77	0,000215	1
200046_at	DAD1: defender against cell death 1	NM_001344	3619,69	2215,97	-1,63	0,000109	1

200617_at	KIAA0152: KIAA0152	NM_014730	3894	2046,49	-1,9	0,000341	1
200644_at	MARCKSL1: MARCKS-like 1	NM_023009	8499,56	4469,02	-1,9	0,000819	1
200694_s_at	DDX24: DEAD (Asp-Glu-Ala-Asp) box polypeptide 24	NM_020414	3485,62	1343,44	-2,59	2,65E-07	0,01444758
200697_at	HK1: hexokinase 1	NM_000188	2498,4	1338,17	-1,87	0,000109	1
200708_at	GOT2: glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)	NM_002080	1317,36	755,8	-1,74	0,000661	1
200842_s_at	EPRS: glutamyl-prolyl-tRNA synthetase	AI475965	765,78	1485,19	1,94	3,42E-05	1
200844_s_at	PRDX6: peroxiredoxin 6	BE869583	2189,52	3774,25	1,72	0,000747	1
200845_s_at	PRDX6: peroxiredoxin 6	NM_004905	1474,45	2627,43	1,78	7,91E-05	1
200861_at	CNOT1: CCR4-NOT transcription complex, subunit 1	NM_016284	409,21	256,59	-1,59	0,000334	1
200882_s_at	PSMD4: proteasome (prosome, macropain) 26S subunit, non-ATPase, 4	NM_002810	1300,23	2059,8	1,58	0,000454	1
200884_at	CKB: creatine kinase, brain	NM_001823	3336,49	1088,3	-3,07	0,000262	1
200894_s_at	FKBP4: FK506 binding protein 4, 59kDa	AA894574	696,56	365,47	-1,91	0,000828	1
200895_s_at	FKBP4: FK506 binding protein 4, 59kDa	NM_002014	2240,59	1256,95	-1,78	0,000845	1
200942_s_at	HSBP1: heat shock factor binding protein 1	NM_001537	3959,15	2336,02	-1,69	5,68E-05	1
200948_at	MLF2: myeloid leukemia factor 2	NM_005439	2469,78	1435,83	-1,72	0,000194	1
201044_x_at	DUSP1: dual specificity phosphatase 1	AA530892	1181,03	270,29	-4,37	0,000827	1
201067_at	PSMC2: proteasome (prosome, macropain) 26S subunit, ATPase, 2	BF215487	1011,78	551,53	-1,83	0,00073	1
201136_at	PLP2: proteolipid protein 2 (colonic epithelium-enriched)	NM_002668	143,07	726,33	5,08	2,73E-05	1
201147_s_at	TIMP3: TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)	BF347089	1948,86	797,87	-2,44	0,000408	1
201148_s_at	TIMP3: TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)	AW338933	797,49	293,97	-2,71	0,00023	1
201149_s_at	TIMP3: TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)	U67195	984,64	374,48	-2,63	0,000483	1
201150_s_at	TIMP3: TIMP metalloproteinase inhibitor 3 (Sorsby fundus dystrophy, pseudoinflammatory)	NM_000362	7877,68	3015,14	-2,61	0,000153	1
201174_s_at	TERF2IP: telomeric repeat binding factor 2, interacting protein GNAI3: guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 3	NM_018975	1915,79	1037,16	-1,85	0,000264	1
201180_s_at	RRBP1: ribosome binding protein 1 homolog 180kDa (dog)	J03198	829,63	1323,32	1,6	0,00029	1
201204_s_at	RRBP1: ribosome binding protein 1 homolog 180kDa (dog)	AA706065	6592,29	3258,49	-2,02	0,000288	1
201206_s_at	RRBP1: ribosome binding protein 1 homolog 180kDa (dog)	NM_004587	1254,52	600,67	-2,09	0,000223	1

201236_s_at	BTG2: BTG family, member 2	NM_006763	2639,9	946,39	-2,79	0,000849	1
201263_at	TARS: threonyl-tRNA synthetase	NM_003191	894,59	1484,94	1,66	0,000738	1
201291_s_at	TOP2A: topoisomerase (DNA) II alpha 170kDa	AU159942	18,53	352,24	19,01	6,98E-05	1
201292_at	TOP2A: topoisomerase (DNA) II alpha 170kDa	AL561834	29,71	389,9	13,12	0,000187	1
201304_at	NDUFA5: NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa	NM_005000	4644,66	2600,03	-1,79	0,000409	1
201346_at	ADIPOR2: adiponectin receptor 2	NM_024551	1339,08	618,88	-2,16	6,74E-05	1
201407_s_at	PPP1CB: protein phosphatase 1, catalytic subunit, beta isoform	AI186712	867,19	1577,85	1,82	3,66E-05	1
201418_s_at	SOX4: SRY (sex determining region Y)-box 4	NM_003107	234,25	1060,35	4,53	0,000468	1
201477_s_at	RRM1: ribonucleotide reductase M1 polypeptide	NM_001033	380,03	1033,09	2,72	0,000186	1
201637_s_at	FXR1: fragile X mental retardation, autosomal homolog 1	NM_005087	1063,52	1777,88	1,67	3,01E-05	1
201650_at	KRT19: keratin 19	NM_002276	111,93	554,92	4,96	0,00053	1
201685_s_at	TOX4: TOX high mobility group box family member 4	NM_014828	747,96	390,45	-1,92	2,76E-06	0,15084441
201816_s_at	GBAS: glioblastoma amplified sequence	NM_001483	6396,89	3096,33	-2,07	0,000109	1
201890_at	RRM2: ribonucleotide reductase M2 polypeptide	BE966236	56,93	592,75	10,41	0,000472	1
201904_s_at	CTDSPL: CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	BF031714	66,01	240,48	3,64	1,54E-07	0,00841911
201906_s_at	CTDSPL: CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	NM_005808	165,17	351,21	2,13	0,000147	1
201917_s_at	SLC25A36: solute carrier family 25, member 36	AI694452	647,52	1334,05	2,06	0,000633	1
201964_at	SETX: senataxin	N64643	1086,8	554,32	-1,96	0,000585	1
202340_x_at	NR4A1: nuclear receptor subfamily 4, group A, member 1	NM_002135	1478,32	250,01	-5,91	2,87E-05	1
202367_at	CUTL1: cut-like 1, CCAAT displacement protein (Drosophila)	NM_001913	1108,83	540,21	-2,05	4,22E-06	0,23013753
202371_at	TCEAL4: transcription elongation factor A (SII)-like 4	NM_024863	3855,81	2499,36	-1,54	0,000674	1
202421_at	IGSF3: immunoglobulin superfamily, member 3	AB007935	112,96	489,59	4,33	3,39E-05	1
202425_x_at	PPP3CA: protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform	NM_000944	287,57	542,38	1,89	1,43E-05	0,78117435
202461_at	EIF2B2: eukaryotic translation initiation factor 2B, subunit 2 beta, 39kDa	NM_014239	2104,88	1130,99	-1,86	0,000216	1
202484_s_at	MBD2: methyl-CpG binding domain protein 2	AF072242	2874,72	1667,8	-1,72	0,000176	1
202518_at	BCL7B: B-cell CLL/lymphoma 7B	NM_001707	486,78	308,59	-1,58	0,000532	1
202539_s_at	HMGCR: 3-hydroxy-3-methylglutaryl-Coenzyme A reductase	AL518627	1101,87	462,64	-2,38	6,95E-05	1
202540_s_at	HMGCR: 3-hydroxy-3-methylglutaryl-Coenzyme A reductase	NM_000859	703,11	288,46	-2,44	1,63E-05	0,88957246

202619_s_at	PLD2: procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	AI754404	1218,56	2538,4	2,08	0,000227	1
202623_at	EAPP: E2F-associated phosphoprotein	NM_018453	1165,75	764,91	-1,52	0,000841	1
202747_s_at	ITM2A: integral membrane protein 2A	NM_004867	792,94	292,03	-2,72	0,00075	1
202768_at	FOSB: FBJ murine osteosarcoma viral oncogene homolog B	NM_006732	4161,57	584,74	-7,12	0,000313	1
202809_s_at	INTS3: integrator complex subunit 3	NM_023015	1669,88	942,06	-1,77	0,000255	1
202843_at	DNAJB9: DnaJ (Hsp40) homolog, subfamily B, member 9	NM_012328	3373,32	1563,23	-2,16	0,000814	1
202854_at	HPRT1: hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)	NM_000194	1027,86	1795,74	1,75	3,42E-05	1
202863_at	SP100: SP100 nuclear antigen	NM_003113	170,63	544,22	3,19	0,000306	1
202867_s_at	DNAJB12: DnaJ (Hsp40) homolog, subfamily B, member 12	NM_017626	528,61	328,49	-1,61	0,000603	1
203097_s_at	RAPGEF2: Rap guanine nucleotide exchange factor (GEF) 2	NM_014247	2318,8	1197,88	-1,94	0,000347	1
203226_s_at	TSPAN31: tetraspanin 31	AL514076	791,04	382,46	-2,07	0,000292	1
203227_s_at	TSPAN31: tetraspanin 31	NM_005981	1927,02	1161,5	-1,66	0,0005	1
203228_at	PAFAH1B3: platelet-activating factor acetylhydrolase, isoform lb, gamma subunit 29kDa	NM_002573	95,37	246,61	2,59	0,0002	1
203339_at	SLC25A12: solute carrier family 25 (mitochondrial carrier, Aralar), member 12	AI887457	280,05	144,55	-1,94	0,000441	1
203354_s_at	PSD3: pleckstrin and Sec7 domain containing 3	AW117368	42,24	263,74	6,24	2,09E-05	1
203355_s_at	PSD3: pleckstrin and Sec7 domain containing 3	NM_015310	115,74	467,9	4,04	0,000157	1
203367_at	DUSP14: dual specificity phosphatase 14	NM_007026	810,4	332,13	-2,44	1,81E-05	0,98834605
203415_at	PDCD6: programmed cell death 6	NM_013232	2713,15	1429,62	-1,9	0,000123	1
203454_s_at	ATOX1: ATX1 antioxidant protein 1 homolog (yeast)	NM_004045	1488,43	774,67	-1,92	0,000695	1
203611_at	TERF2: telomeric repeat binding factor 2	NM_005652	450,36	260,83	-1,73	0,000431	1
203630_s_at	COG5: component of oligomeric golgi complex 5	NM_006348	994,55	529,53	-1,88	0,000763	1
203717_at	DPP4: dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	NM_001935	19,92	246	12,35	0,000784	1
203733_at	DEXI: dexamethasone-induced transcript	NM_014015	715,8	373,14	-1,92	0,000206	1
203766_s_at	LMOD1: leiomodlin 1 (smooth muscle)	NM_012134	268,95	62,5	-4,3	6,42E-05	1
203786_s_at	TPD52L1: tumor protein D52-like 1	NM_003287	467,41	1280,82	2,74	0,000352	1
203980_at	FABP4: fatty acid binding protein 4, adipocyte	NM_001442	2541,93	448,25	-5,67	0,000135	1
204027_s_at	METTL1: methyltransferase like 1	NM_005371	201,47	81,45	-2,47	0,000239	1
204202_at	IQCE: IQ motif containing E	NM_017604	506,29	185,24	-2,73	3,94E-06	0,21522974
204246_s_at	DCTN3: dynactin 3 (p22)	NM_007234	2177,3	1160,09	-1,88	7,04E-06	0,38415828
204274_at	EBAG9: estrogen receptor binding site associated, antigen, 9	AA812215	2917,91	1820,88	-1,6	0,000611	1

204285_s_at	PMAIP1: phorbol-12-myristate-13-acetate-induced protein 1	AI857639	19,85	245,05	12,35	3,27E-05	1
204368_at	SLCO2A1: solute carrier organic anion transporter family, member 2A1	NM_005630	1506,89	424,22	-3,55	4,23E-05	1
204451_at	FZD1: frizzled homolog 1 (Drosophila)	NM_003505	1792,1	534,96	-3,35	3,54E-05	1
204513_s_at	ELMO1: engulfment and cell motility 1	NM_014800	830,82	152,41	-5,45	8,33E-06	0,45468042
204573_at	CROT: carnitine O-octanoyltransferase	NM_021151	3376,59	926,33	-3,65	4,01E-05	1
204576_s_at	CLUAP1: clusterin associated protein 1	AA207013	313,83	175,14	-1,79	0,000218	1
204675_at	SRD5A1: steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)	NM_001047	20,85	149,28	7,16	0,00013	1
204686_at	IRS1: insulin receptor substrate 1	NM_005544	1710,37	755,31	-2,26	0,000246	1
204724_s_at	COL9A3: collagen, type IX, alpha 3	NM_001853	10407,18	3868,7	-2,69	0,000799	1
204765_at	ARHGEF5: Rho guanine nucleotide exchange factor (GEF) 5	NM_005435	532,56	282,11	-1,89	0,000758	1
204832_s_at	BMPRI1A: bone morphogenetic protein receptor, type IA	NM_004329	741,37	345,67	-2,14	0,000265	1
204881_s_at	UGCG: UDP-glucose ceramide glucosyltransferase	NM_003358	168,76	657,15	3,89	0,000587	1
204932_at	TNFRSF11B: tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)	BF433902	587,22	208,71	-2,81	0,000125	1
204933_s_at	TNFRSF11B: tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)	NM_002546	1025,1	278,02	-3,69	0,000115	1
205012_s_at	HAGH: hydroxyacylglutathione hydrolase	NM_005326	556,56	300,73	-1,85	0,000285	1
205042_at	GNE: glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase	NM_005476	214,61	109,99	-1,95	0,000346	1
205051_s_at	KIT: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	NM_000222	2360,53	829,12	-2,85	0,000727	1
205052_at	AUH: AU RNA binding protein/enoyl-Coenzyme A hydratase	NM_001698	1445,56	782,86	-1,85	0,000425	1
205084_at	BCAP29: B-cell receptor-associated protein 29	NM_018844	1644,68	645,79	-2,55	3,86E-05	1
205094_at	PEX12: peroxisomal biogenesis factor 12	NM_000286	422,04	201,16	-2,1	0,000144	1
205247_at	NOTCH4: Notch homolog 4 (Drosophila)	AI743713	354,05	163,8	-2,16	0,000507	1
205348_s_at	DYNC11I: dynein, cytoplasmic 1, intermediate chain 1	NM_004411	251,5	67,91	-3,7	0,000645	1
205350_at	CRABP1: cellular retinoic acid binding protein 1	NM_004378	1757,89	705,69	-2,49	0,000239	1
205357_s_at	AGTR1: angiotensin II receptor, type 1	NM_000685	201,37	29,72	-6,78	7,55E-05	1
823_at	CX3CL1: chemokine (C-X3-C motif) ligand 1	U84487	244,15	107,99	-2,26	0,000823	1
205554_s_at	DNASE1L3: deoxyribonuclease I-like 3	NM_004944	198,98	40,07	-4,97	5,43E-06	0,29622852
205676_at	CYP27B1: cytochrome P450, family 27, subfamily B, polypeptide 1	NM_000785	681,75	81,92	-8,32	0,000336	1

206208_at	CA4: carbonic anhydrase IV	NM_000717	361,46	53,39	-6,77	3,88E-06	0,21155011
206209_s_at	CA4: carbonic anhydrase IV	NM_000717	552,08	59,25	-9,32	5,5E-06	0,30034542
206214_at	PLA2G7: phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	NM_005084	1288,54	294,22	-4,38	0,000246	1
206355_at	GNAL: guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type	R20102	413,35	75,03	-5,51	2,15E-05	1
206396_at	SLC1A1: solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	NM_004170	245,46	47,14	-5,21	1,16E-05	0,6346978
206517_at	CDH16: cadherin 16, KSP-cadherin	NM_004062	619,48	63,33	-9,78	1,82E-07	0,00995896
206656_s_at	C20orf3: chromosome 20 open reading frame 3	BC000353	4520,93	2464,11	-1,83	0,00019	1
206790_s_at	NDUFB1: NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	NM_004545	8346,85	5310	-1,57	0,000363	1
206806_at	DGKI: diacylglycerol kinase, iota	NM_004717	741,91	142,97	-5,19	0,000107	1
207826_s_at	ID3: inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	NM_002167	9357,11	3478,19	-2,69	0,00042	1
207974_s_at	SKP1A: S-phase kinase-associated protein 1A (p19A)	NM_006930	7056,61	4408,84	-1,6	2,15E-05	1
207980_s_at	CITED2: Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	NM_006079	3912,99	980,87	-3,99	0,000206	1
208140_s_at	LRR48: leucine rich repeat containing 48	NM_031294	193,28	85,1	-2,27	0,00052	1
208374_s_at	CAPZA1: capping protein (actin filament) muscle Z-line, alpha 1	NM_006135	2096,8	3330,93	1,59	0,000266	1
208785_s_at	MAP1LC3B: microtubule-associated protein 1 light chain 3 beta	BE893893	921,59	572,12	-1,61	0,000111	1
208786_s_at	MAP1LC3B: microtubule-associated protein 1 light chain 3 beta	AF183417	2761,04	1435,7	-1,92	3,24E-05	1
208809_s_at	C6orf62: chromosome 6 open reading frame 62	AL136632	1809,83	2800,79	1,55	0,00024	1
208869_s_at	GABARAPL1: GABA(A) receptor-associated protein like 1	AF087847	380,32	244,82	-1,55	0,000505	1
208928_at	POR: P450 (cytochrome) oxidoreductase	AF258341	3060,85	814,76	-3,76	3,19E-05	1
209006_s_at	C1orf63: chromosome 1 open reading frame 63	AF247168	338,32	846,91	2,5	0,00063	1
209029_at	COPS7A: COP9 constitutive photomorphogenic homolog subunit 7A (Arabidopsis)	AF193844	665,74	391,34	-1,7	0,000121	1
209046_s_at	GABARAPL2: GABA(A) receptor-associated protein-like 2	AB030710	5249,08	2924,06	-1,8	0,000171	1
209049_s_at	ZMYND8: zinc finger, MYND-type containing 8	BC001004	465,55	916,21	1,97	1,32E-05	0,71748059
209147_s_at	PPAP2A: phosphatidic acid phosphatase type 2A	AB000888	2010,47	924,72	-2,17	0,000633	1
209149_s_at	TM9SF1: transmembrane 9 superfamily member 1	BE899402	552,49	331,65	-1,67	0,000251	1
209229_s_at	SAPS1: SAPS domain family, member 1	BC002799	442,88	266,19	-1,66	0,000813	1

209242_at	PEG3: paternally expressed 3	AL042588	1330,36	355,4	-3,74	0,000492	1
209311_at	BCL2L2: BCL2-like 2	D87461	828,98	323,18	-2,57	1,64E-06	0,0895375
209325_s_at	RGS16: regulator of G-protein signaling 16	U94829	1335,51	350,76	-3,81	0,000301	1
209357_at	CITED2: Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	AF109161	5353,85	1763,86	-3,04	0,000474	1
209418_s_at	THOC5: THO complex 5	BC003615	109,04	230,4	2,11	0,000638	1
209531_at	GSTZ1: glutathione transferase zeta 1 (maleylacetoacetate isomerase)	BC001453	491,89	256,57	-1,92	0,000621	1
209579_s_at	MBD4: methyl-CpG binding domain protein 4	AL556619	690,25	1244,42	1,8	0,000156	1
209623_at	MCCC2: methylcrotonoyl-Coenzyme A carboxylase 2 (beta)	AW439494	2181,99	925,06	-2,36	0,000609	1
209682_at	CBLB: Cas-Br-M (murine) ecotropic retroviral transforming sequence b	U26710	296,09	899,7	3,04	0,000586	1
209735_at	ABCG2: ATP-binding cassette, sub-family G (WHITE), member 2	AF098951	179,59	48,43	-3,71	1,47E-05	0,80374748
209773_s_at	RRM2: ribonucleotide reductase M2 polypeptide	BC001886	28,84	310,28	10,76	0,000355	1
209814_at	ZNF330: zinc finger protein 330	BC004421	1340,77	764,86	-1,75	7,44E-05	1
209897_s_at	SLIT2: slit homolog 2 (Drosophila)	AF055585	38,25	150,97	3,95	0,000783	1
209917_s_at	TP53AP1: TP53 activated protein 1	BC002709	802,4	257,7	-3,11	9,45E-06	0,51571359
209945_s_at	GSK3B: glycogen synthase kinase 3 beta	BC000251	233,91	541,54	2,32	3,95E-05	1
209998_at	PIGO: phosphatidylinositol glycan anchor biosynthesis, class O	BC001030	323,89	179,72	-1,8	0,00033	1
210149_s_at	ATP5H: ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	AF061735	9285,36	6150,01	-1,51	0,000473	1
210241_s_at	TP53AP1: TP53 activated protein 1	AB007458	821,32	307,59	-2,67	2,36E-05	1
210529_s_at	FAM115A /// FAM115B: family with sequence similarity 115, member A /// family with sequence similarity 115, member B	BC000609	333,86	128,23	-2,6	0,000277	1
210534_s_at	EPB9: B9 protein	BC002944	621,69	223,42	-2,78	0,00029	1
210788_s_at	DHRS7: dehydrogenase/reductase (SDR family) member 7	AF126782	3694,16	1776,71	-2,08	3,58E-05	1
210817_s_at	CALCOCO2: calcium binding and coiled-coil domain 2	BC004130	1013,69	572,65	-1,77	0,000264	1
210825_s_at	PEBP1: phosphatidylethanolamine binding protein 1	AF130103	24021	13931,2	-1,72	0,00031	1
210840_s_at	IQGAP1: IQ motif containing GTPase activating protein 1	D29640	738,47	1682,43	2,28	0,000159	1
210886_x_at	TP53AP1: TP53 activated protein 1	AB007457	944,02	370,02	-2,55	2,02E-05	1
211034_s_at	FLJ30092: AF-1 specific protein phosphatase	BC006270	833,28	403,15	-2,07	0,000701	1
211143_x_at	NR4A1: nuclear receptor subfamily 4, group A, member 1	D49728	230,67	76,66	-3,01	2,67E-05	1
211217_s_at	KCNQ1: potassium voltage-gated channel, KQT-like subfamily,	AF051426	480,89	179,49	-2,68	0,000458	1

211453_s_at	member 1	M77198	233,52	86,97	-2,69	0,000161	1
211941_s_at	AKT2: v-akt murine thymoma viral oncogene homolog 2	BE969671	12787,54	6869,6	-1,86	0,000405	1
211944_at	PEBP1: phosphatidylethanolamine binding protein 1	BE729523	103,45	275,01	2,66	0,000514	1
212041_at	BAT2D1: BAT2 domain containing 1	AL566172	3284,44	1687,22	-1,95	5,36E-05	1
212114_at	ATP6V0D1: ATPase, H+ transporting, lysosomal 38kDa, V0 subunit d1	BE967207	2087,23	1136,59	-1,84	0,00017	1
212228_s_at	LOC552889: hypothetical LOC552889	AC004382	1645,68	860,29	-1,91	0,000845	1
212299_at	COQ9: coenzyme Q9 homolog (S. cerevisiae)	AL117502	610,44	368,55	-1,66	0,000277	1
212310_at	NEK9: NIMA (never in mitosis gene a)- related kinase 9	D87742	297,28	135,39	-2,2	0,000554	1
212508_at	MIA3: melanoma inhibitory activity family, member 3	AK024029	2794,65	1490,91	-1,87	0,000263	1
212564_at	MOAP1: modulator of apoptosis 1	AA523921	372,91	167,46	-2,23	0,000175	1
212591_at	KCTD2: potassium channel tetramerisation domain containing 2	AA887480	556,89	955,66	1,72	0,000216	1
212595_s_at	RBM34: RNA binding motif protein 34	AL534321	2763,89	1501,05	-1,84	0,000817	1
212596_s_at	DAZAP2: DAZ associated protein 2	AJ010070	263,13	414,39	1,57	0,000374	1
212680_x_at	HMG2L1: high-mobility group protein 2-like 1	BE305165	355,94	1020,74	2,87	0,000217	1
212685_s_at	PPP1R14B: protein phosphatase 1, regulatory (inhibitor) subunit 14B	AI608789	2109,65	1259,84	-1,67	1,73E-05	0,94509202
212708_at	TBL2: transducin (beta)-like 2	AV721987	1515,31	753,54	-2,01	5,39E-05	1
212814_at	MSL-1: male-specific lethal-1 homolog	AB020635	1341,67	341,49	-3,93	3,08E-06	0,16814665
212840_at	KIAA0828: adenosylhomocysteinase 3	BG339560	168,64	413,4	2,45	0,00016	1
212852_s_at	UBXD7: UBX domain containing 7	AL538601	2275,04	3816,14	1,68	3,55E-05	1
212880_at	TROVE2: TROVE domain family, member 2	AB011113	510,58	242,15	-2,11	0,00016	1
212897_at	WDR7: WD repeat domain 7	AI738802	82,32	207,32	2,52	5,64E-05	1
212899_at	CDC2L6: cell division cycle 2-like 6 (CDK8-like)	AB028951	124,93	243,27	1,95	0,00062	1
213099_at	CDC2L6: cell division cycle 2-like 6 (CDK8-like)	AB018302	249,94	139,86	-1,79	0,000307	1
213185_at	ANGEL1: angel homolog 1 (Drosophila)	AI758896	466,67	203,45	-2,29	0,000162	1
213190_at	KIAA0556: KIAA0556	R61519	357,06	188,55	-1,89	7,72E-05	1
213195_at	COG7: component of oligomeric golgi complex 7	AI625844	378,57	155,19	-2,44	0,000528	1
213226_at	LOC201229: hypothetical protein LOC201229	AI346350	12,42	126,51	10,19	0,000335	1
213234_at	CCNA2: cyclin A2	AB040900	324,06	55,81	-5,81	1,61E-06	0,08792191
213246_at	KIAA1467: KIAA1467	AI346504	1281,51	678,36	-1,89	0,000571	1
	C14orf109: chromosome 14 open reading frame 109						

213333_at	MDH2: malate dehydrogenase 2, NAD (mitochondrial)	AL520774	301,65	165,7	-1,82	0,000627	1
213351_s_at	TMCC1: transmembrane and coiled-coil domain family 1	AB018322	180,37	299,76	1,66	0,000301	1
213392_at	IQCK: IQ motif containing K PHLPL: PH domain and leucine rich repeat protein phosphatase-like	AW070229	1149,52	560,84	-2,05	9,33E-05	1
213407_at		AB023148	511,95	184,82	-2,77	0,000172	1
213444_at	LOC643641: hypothetical protein LOC643641	AB011115	551,54	215,82	-2,56	4,43E-05	1
213488_at	SNED1: sushi, nidogen and EGF-like domains 1	N73970	292,64	123,84	-2,36	0,000568	1
213508_at	C14orf147: chromosome 14 open reading frame 147 EIF4E2: eukaryotic translation initiation factor 4E family member 2	AA142942	981,64	467,53	-2,1	5,44E-06	0,29680405
213571_s_at		BF516289	268,45	438,34	1,63	0,000696	1
213578_at	BMPR1A: bone morphogenetic protein receptor, type IA	AI678679	1234,3	672,51	-1,84	0,000338	1
213637_at	Transcribed locus	BE503392	82,07	208,85	2,54	3,37E-05	1
213664_at	SLC1A1: solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	AW235061	1954,68	474,69	-4,12	0,000545	1
213742_at	SFRS11: splicing factor, arginine/serine-rich 11	AW241752	175,28	492,56	2,81	0,000621	1
213822_s_at	UBE3B: ubiquitin protein ligase E3B	BE856776	281,96	83,07	-3,39	3,06E-07	0,01670057
213861_s_at	FAM119B: family with sequence similarity 119, member B	N67741	308,34	108,77	-2,83	2,09E-05	1
213924_at	MPPE1: Metallophosphoesterase 1	BF476502	876,35	237,41	-3,69	2,73E-05	1
214071_at	MPPE1: Metallophosphoesterase 1	AI082827	316,66	103,38	-3,06	0,00058	1
214180_at	MAN1C1: mannosidase, alpha, class 1C, member 1	AW340588	378,57	134,47	-2,82	0,00078	1
214722_at	NOTCH2NL: Notch homolog 2 (Drosophila) N-terminal like	AW516297	1063,78	2309,57	2,17	0,000722	1
214733_s_at	YIPF1: Yip1 domain family, member 1	AL031427	969,47	396,36	-2,45	2,05E-05	1
214746_s_at	ZNF467: zinc finger protein 467	BE549732	237,75	84,28	-2,82	0,000115	1
215130_s_at	IQCK: IQ motif containing K	AC002550	202,86	90,97	-2,23	5,99E-05	1
217188_s_at	C14orf1: chromosome 14 open reading frame 1	AC007182	384,36	196,12	-1,96	0,000672	1
217216_x_at	MLH3: mutL homolog 3 (E. coli)	AC006530	176,96	72,73	-2,43	2,58E-05	1
217523_at	CD44: CD44 molecule (Indian blood group) MBTPS1: membrane-bound transcription factor peptidase, site 1	AV700298	229,84	1123,42	4,89	0,000617	1
217543_s_at		BE890314	520,22	288,61	-1,8	0,00012	1
217726_at	COPZ1: coatomer protein complex, subunit zeta 1	NM_016057	2434,43	1389,23	-1,75	8,22E-05	1
217732_s_at	ITM2B: integral membrane protein 2B	AF092128	15848,07	9797,19	-1,62	0,000325	1
217751_at	GSTK1: glutathione S-transferase kappa 1	NM_015917	3107,11	1761,51	-1,76	0,000611	1
217853_at	TNS3: tensin 3	NM_022748	2341,07	1033,67	-2,26	0,000459	1
217957_at	C16orf80: chromosome 16 open reading frame 80	NM_013242	1085,02	643,12	-1,69	0,00012	1

217966_s_at	FAM129A: family with sequence similarity 129, member A	NM_022083	22	207,26	9,42	0,000194	1
217968_at	TSSC1: tumor suppressing subtransferable candidate 1	NM_003310	744,58	352,84	-2,11	0,000396	1
217990_at	GMPPR2: guanosine monophosphate reductase 2	NM_016576	1244,98	670,54	-1,86	0,000325	1
218034_at	FIS1: fission 1 (mitochondrial outer membrane) homolog (S. cerevisiae)	NM_016068	4265,74	2582,87	-1,65	0,00054	1
218039_at	NUSAP1: nucleolar and spindle associated protein 1	NM_016359	36,63	385,87	10,53	0,000407	1
218177_at	CHMP1B: chromatin modifying protein 1B	AA293502	1370,57	363,97	-3,77	2,34E-05	1
218178_s_at	CHMP1B: chromatin modifying protein 1B	NM_020412	3085,35	1239,03	-2,49	0,000446	1
218200_s_at	NDUFB2: NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	NM_004546	14012,75	6838,2	-2,05	1,83E-05	1
218201_at	NDUFB2: NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	NM_004546	3873,71	2113,78	-1,83	0,000616	1
218241_at	GOLGA5: golgi autoantigen, golgin subfamily a, 5	NM_005113	1101,72	695,52	-1,58	3,28E-05	1
218273_s_at	PPM2C: protein phosphatase 2C, magnesium-dependent, catalytic subunit	NM_018444	95,61	273,29	2,86	9,86E-05	1
218292_s_at	PRKAG2: protein kinase, AMP-activated, gamma 2 non-catalytic subunit	NM_016203	181,42	74,22	-2,44	9,06E-05	1
218300_at	C16orf53: chromosome 16 open reading frame 53	NM_024516	320,35	171,92	-1,86	0,000164	1
218328_at	COQ4: coenzyme Q4 homolog (S. cerevisiae)	NM_016035	1405,38	553,8	-2,54	0,000353	1
218349_s_at	ZWILCH: Zwilch, kinetochore associated, homolog (Drosophila)	NM_017975	22,29	138,75	6,22	0,000625	1
218450_at	HEBP1: heme binding protein 1	NM_015987	6236,33	2331,65	-2,67	1,42E-05	0,77346231
218487_at	ALAD: aminolevulinate, delta-, dehydratase	BC000977	811,75	411,81	-1,97	0,000214	1
218504_at	FAHD2A: fumarylacetoacetate hydrolase domain containing 2A	NM_016044	570,77	218,26	-2,62	0,000287	1
218540_at	THTPA: thiamine triphosphatase	NM_024328	496,96	220,44	-2,25	0,000178	1
218559_s_at	MAFB: v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	NM_005461	2056,44	726,55	-2,83	0,000127	1
218584_at	TECT1: tectonic 1	NM_024549	524,55	230,74	-2,27	0,00044	1
218601_at	URG4: up-regulated gene 4	NM_017920	467,79	281,48	-1,66	0,000679	1
218667_at	PJA1: praja 1	NM_022368	985,72	474,41	-2,08	0,000102	1
218723_s_at	C13orf15: chromosome 13 open reading frame 15	NM_014059	3891,36	1305,13	-2,98	0,000109	1
	LOC727901 /// VWA1: von Willebrand factor A domain containing 1 /// similar to von Willebrand factor A domain-related protein isoform 1						
218731_s_at	TAPBPL: TAP binding protein-like	NM_022834	375,84	179,35	-2,1	0,000121	1
218746_at		NM_018009	615,94	261,18	-2,36	0,000722	1

218747_s_at	TAPBPL: TAP binding protein-like	NM_018009	539,41	254,22	-2,12	0,000717	1
218750_at	JOSD3: Josephin domain containing 3	NM_024116	58,79	184,06	3,13	0,000154	1
218773_s_at	MSRB2: methionine sulfoxide reductase B2	NM_012228	3473,29	2115,8	-1,64	0,000651	1
218866_s_at	POLR3K: polymerase (RNA) III (DNA directed) polypeptide K, 12.3 kDa	NM_016310	451,11	272,08	-1,66	0,000646	1
218918_at	MAN1C1: mannosidase, alpha, class 1C, member 1	NM_020379	943,27	307,2	-3,07	0,000802	1
219015_s_at	ALG13: asparagine-linked glycosylation 13 homolog (S. cerevisiae)	NM_018466	533,05	223,72	-2,38	0,000639	1
219064_at	ITIH5: inter-alpha (globulin) inhibitor H5	NM_030569	219,51	64,23	-3,42	7,06E-05	1
219091_s_at	MMRN2: multimerin 2	NM_024756	654,86	292,98	-2,24	0,0004	1
219163_at	ZNF562: zinc finger protein 562	NM_017656	272,3	438,36	1,61	0,000714	1
219182_at	FLJ22167: hypothetical protein FLJ22167	NM_024533	285,59	100,28	-2,85	0,000418	1
219238_at	PIGV: phosphatidylinositol glycan anchor biosynthesis, class V	NM_017837	503,11	213,54	-2,36	1,11E-06	0,06059004
219259_at	SEMA4A: sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4A	NM_022367	261,21	83,07	-3,14	0,000101	1
219286_s_at	RBM15: RNA binding motif protein 15	NM_022768	263,95	436,31	1,65	0,000637	1
219321_at	MPP5: membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	NM_022474	593,1	280,32	-2,12	0,000494	1
219436_s_at	EMCN: endomucin	NM_016242	380,74	128	-2,97	0,00038	1
219455_at	FLJ21062: hypothetical protein FLJ21062	NM_024788	245,85	80,35	-3,06	0,000128	1
219798_s_at	MEPCE: methylphosphate capping enzyme	NM_019606	1702,38	942,19	-1,81	0,000434	1
219816_s_at	RBM23: RNA binding motif protein 23	NM_018107	1252,58	606,31	-2,07	1,57E-05	0,85439904
219825_at	CYP26B1: cytochrome P450, family 26, subfamily B, polypeptide 1	NM_019885	130,3	27,84	-4,68	0,000604	1
219862_s_at	NARF: nuclear prelamin A recognition factor	NM_012336	286,56	498,7	1,74	2,9E-06	0,15854512
219877_at	ZMAT4: zinc finger, matrin type 4	NM_024645	278,5	42,89	-6,49	3,69E-06	0,20157656
219929_s_at	ZFYVE21: zinc finger, FYVE domain containing 21	NM_024071	546,3	193,07	-2,83	6,6E-06	0,36008884
219956_at	GALNT6: UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)	NM_007210	10,09	115,44	11,44	0,000629	1
220108_at	GNA14: guanine nucleotide binding protein (G protein), alpha 14	NM_004297	396,71	115,98	-3,42	0,000328	1
220173_at	C14orf45: chromosome 14 open reading frame 45	NM_025057	497,87	186,46	-2,67	6,8E-05	1
220203_at	BMP8A: bone morphogenetic protein 8a	NM_024732	847,41	148,01	-5,73	0,000162	1

220204_s_at	BMP8A: bone morphogenetic protein 8a	NM_024732	1212,37	175	-6,93	0,000115	1
220751_s_at	C5orf4: chromosome 5 open reading frame 4	NM_016348	520,3	158,4	-3,28	0,000523	1
221008_s_at	AGXT2L1: alanine-glyoxylate aminotransferase 2-like 1	NM_031279	171,85	8,97	-19,16	0,000629	1
221031_s_at	APOLD1: apolipoprotein L domain containing 1	NM_030817	1917,63	658,37	-2,91	5,24E-05	1
221046_s_at	GTPBP8: GTP-binding protein 8 (putative)	NM_014170	257,91	507	1,97	8,38E-05	1
221447_s_at	GLT8D2: glycosyltransferase 8 domain containing 2	NM_031302	1492,45	155,34	-9,61	4,98E-07	0,02719047
221449_s_at	ITFG1: integrin alpha FG-GAP repeat containing 1	NM_030790	1698,86	773,43	-2,2	3,11E-06	0,16967494
221529_s_at	PLVAP: plasmalemma vesicle associated protein	AF326591	3521,99	1358,07	-2,59	0,000229	1
221619_s_at	MTCH1: mitochondrial carrier homolog 1 (C. elegans)	AF189289	16805,97	9353,26	-1,8	7,11E-05	1
221686_s_at	RECQL5: RecQ protein-like 5	AL136869	488,42	189,47	-2,58	1,1E-05	0,5980379
221776_s_at	BRD7: bromodomain containing 7	AI885109	841,63	509,61	-1,65	0,000225	1
221820_s_at	MYST1: MYST histone acetyltransferase 1	AK024102	1203,37	727,7	-1,65	0,00078	1
221846_s_at	CASKIN2: CASK interacting protein 2	AI970096	283,82	142,31	-1,99	0,000361	1
222056_s_at	FAHD2A: fumarylacetoacetate hydrolase domain containing 2A	AA723370	1878,05	800,03	-2,35	0,000212	1
222068_s_at	LRR50: leucine rich repeat containing 50	AW663632	187,93	54,04	-3,48	3,6E-05	1
222386_s_at	COPZ1: coatomer protein complex, subunit zeta 1	AB047848	2748,59	1658,69	-1,66	8,47E-05	1
222445_at	SLC39A9: solute carrier family 39 (zinc transporter), member 9	AK025631	1379,9	802,33	-1,72	2,24E-05	1
222608_s_at	ANLN: anillin, actin binding protein	AK023208	7,94	198,13	24,95	0,000767	1
222664_at	KCTD15: potassium channel tetramerisation domain containing 15	AI808448	263,13	128,3	-2,05	0,000567	1
222670_s_at	MAFB: v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	AW135013	1730,41	678,73	-2,55	0,000289	1
222728_s_at	JOSD3: Josephin domain containing 3	AF275800	286,07	871,64	3,05	1,82E-06	0,09960317
222742_s_at	RABL5: RAB, member RAS oncogene family-like 5	AW026449	1619,64	769,71	-2,1	0,000409	1
222778_s_at	WHSC1: Wolf-Hirschhorn syndrome candidate 1	AW024870	179,79	63,66	-2,82	0,000768	1
222785_x_at	C11orf1: chromosome 11 open reading frame 1	AJ250229	894,88	419,86	-2,13	0,00021	1
222798_at	PTER: phosphotriesterase related	BF112019	303,4	678,68	2,24	0,000766	1
222808_at	ALG13: asparagine-linked glycosylation 13 homolog (S. cerevisiae)	BC005336	1447,01	672,31	-2,15	0,000126	1
222833_at	AYTL1: acyltransferase like 1	AU154202	492,25	199,24	-2,47	0,000222	1
222835_at	THSD4: thrombospondin, type I, domain containing 4	BG163478	53,3	203,16	3,81	0,000681	1
222885_at	EMCN: endomucin	AF205940	373,82	158,3	-2,36	0,000702	1
222983_s_at	PAIP2: poly(A) binding protein interacting protein 2	BC001716	2459,48	1357,16	-1,81	0,000166	1

222984_at	PAIP2: poly(A) binding protein interacting protein 2	AF151052	3169,25	1898,4	-1,67	0,000505	1
222995_s_at	RHBD2: rhomboid domain containing 2	AF226732	3033,65	1380,75	-2,2	0,000465	1
223009_at	C11orf59: chromosome 11 open reading frame 59	BC001706	3360,76	1953,74	-1,72	5,57E-05	1
223053_x_at	SSU72: SSU72 RNA polymerase II CTD phosphatase homolog (S. cerevisiae)	AF277178	874,53	571,42	-1,53	4,85E-05	1
223114_at	COQ5: coenzyme Q5 homolog, methyltransferase (S. cerevisiae)	BC004916	1581,24	920,86	-1,72	0,000459	1
223148_at	PIGS: phosphatidylinositol glycan anchor biosynthesis, class S	BC001319	274,17	96,87	-2,83	6,05E-06	0,330038
223162_s_at	KIAA1147: KIAA1147	AF1116707	1027,42	492,8	-2,08	0,000299	1
223259_at	ORMDL3: ORM1-like 3 (S. cerevisiae)	BC000638	1503,33	689,51	-2,18	0,000339	1
223367_at	WBSR18: Williams Beuren syndrome chromosome region 18	BC005056	807,39	399,79	-2,02	8,53E-05	1
223393_s_at	TSHZ3: teashirt zinc finger homeobox 3	AL136805	266,49	105,68	-2,52	0,00065	1
223459_s_at	C1orf56: chromosome 1 open reading frame 56	BE222214	422,37	184,82	-2,29	0,000271	1
223577_x_at	PRO1073: PRO1073 protein	AA827878	306,99	791,55	2,58	0,000564	1
223773_s_at	C1orf79: chromosome 1 open reading frame 79	AF277181	153	421,6	2,76	0,000189	1
223776_x_at	TINF2: TERF1 (TRF1)-interacting nuclear factor 2	BC005030	618,48	401,51	-1,54	0,00047	1
223887_at	GPR132: G protein-coupled receptor 132	BC004555	126,7	20,62	-6,14	0,000722	1
224179_s_at	MIOX: myo-inositol oxygenase	AF230095	207,6	46,13	-4,5	3,31E-05	1
224415_s_at	HINT2: histidine triad nucleotide binding protein 2	AF356515	1516,07	803,59	-1,89	0,000428	1
224445_s_at	ZFYVE21: zinc finger, FYVE domain containing 21	BC005999	1499,9	677,47	-2,21	3,54E-05	1
224559_at	MALAT1: metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)	AF001540	1604,39	5189,94	3,23	0,000321	1
224573_at	MGC71993: similar to DNA segment, Chr 11, Brigham & Womens Genetics 0434 expressed	BE744389	7441,52	4749,95	-1,57	0,00033	1
224578_at	RCC2: regulator of chromosome condensation 2	AB040903	294,37	667,89	2,27	0,000301	1
224598_at	MGAT4B: mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme B	BF570193	597,27	1244,56	2,08	0,000139	1
224628_at	C2orf30: chromosome 2 open reading frame 30	AF131743	3437,69	2177,65	-1,58	0,000708	1
224630_at	C2orf30: chromosome 2 open reading frame 30	AK001913	1118,02	583,35	-1,92	0,000119	1
224688_at	C7orf42: chromosome 7 open reading frame 42	BE962299	1294,79	771,26	-1,68	3,75E-05	1
224698_at	FAM62B: family with sequence similarity 62 (C2 domain containing) member B	AB033054	1571,99	895,56	-1,76	7,48E-05	1
224732_at	CTF8: chromosome transmission fidelity factor 8 homolog (S. cerevisiae)	AI309784	1162,45	617,03	-1,88	8,2E-05	1

224759_s_at	C12orf23: chromosome 12 open reading frame 23	AK001731	2921,48	1650,62	-1,77	0,000343	1
224809_x_at	TINF2: TERF1 (TRF1)-interacting nuclear factor 2	AK023166	640,7	417,13	-1,54	0,000472	1
224879_at	C9orf123: chromosome 9 open reading frame 123	BF315994	1785,17	999,98	-1,79	0,000103	1
224887_at	GNPTG: N-acetylglucosamine-1-phosphate transferase, gamma subunit	AF302786	754,54	462,82	-1,63	0,0002	1
225017_at	CCDC14: coiled-coil domain containing 14	AK022954	169,79	399,73	2,35	0,000824	1
225074_at	RAB2B: RAB2B, member RAS oncogene family	AA531016	544,96	335,05	-1,63	0,000698	1
225096_at	C17orf79: chromosome 17 open reading frame 79	AJ272196	2813,46	1230,41	-2,29	0,000586	1
225134_at	SPRYD3: SPRY domain containing 3	AF131774	852,31	449,35	-1,9	4,1E-05	1
225183_at	C16orf72: chromosome 16 open reading frame 72	BG495327	1224,64	807,02	-1,52	0,000202	1
225198_at	VAPA: VAMP (vesicle-associated membrane protein)-associated protein A, 33kDa	AL571942	1294,21	716,42	-1,81	0,000585	1
225260_s_at	MRPL32: mitochondrial ribosomal protein L32	AL551823	3042,63	1955,19	-1,56	0,000482	1
225283_at	ARRDC4: arrestin domain containing 4	AV701177	1815,06	516,39	-3,51	0,000186	1
225287_s_at	TMEM55B: transmembrane protein 55B	AI992151	364,5	211,68	-1,72	0,000639	1
225415_at	DTX3L: dxtex 3-like (Drosophila)	AA577672	585,51	1201,02	2,05	0,000403	1
225426_at	PPP6C: protein phosphatase 6, catalytic subunit	AW195360	889,84	475,03	-1,87	6,74E-05	1
225446_at	BRWD1: bromodomain and WD repeat domain containing 1	AI638279	315,19	169,79	-1,86	0,000725	1
225471_s_at	AKT2: v-akt murine thymoma viral oncogene homolog 2	BE734905	2900,83	1051,99	-2,76	1,85E-05	1
225484_at	TSGA14: testis specific, 14	AW157525	270,17	109,23	-2,47	0,000269	1
225558_at	GIT2: G protein-coupled receptor kinase interactor 2	R38084	734,53	488,76	-1,5	0,000564	1
225589_at	SH3RF1: SH3 domain containing ring finger 1	AB040927	1004,15	451,68	-2,22	8,91E-06	0,48586385
225677_at	BCAP29: B-cell receptor-associated protein 29	AW152589	2714,91	1212,31	-2,24	0,000153	1
225779_at	SLC27A4: solute carrier family 27 (fatty acid transporter), member 4	AK000722	1528,34	420,88	-3,63	0,000299	1
225798_at	JAZF1: JAZF zinc finger 1	AL047908	1613,46	412,2	-3,91	0,000261	1
225800_at	JAZF1: JAZF zinc finger 1	AI990891	453,01	152,77	-2,97	0,000387	1
225804_at	CYB5D2: cytochrome b5 domain containing 2	BE044480	694,68	205,05	-3,39	0,000107	1
225812_at	LOC619208: hypothetical protein LOC619208	N36759	520,85	256,38	-2,03	0,000518	1
225982_at	UBTF: upstream binding transcription factor, RNA polymerase I	BG341575	409,61	233,28	-1,76	5,6E-05	1
226028_at	ROBO4: roundabout homolog 4, magic roundabout (Drosophila)	AA156022	534,22	233,97	-2,28	0,000799	1
226092_at	MPP5: membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	BF115203	1654,13	753,17	-2,2	0,00032	1

226095_s_at	ATXN1L: ataxin 1-like	AW138861	463,36	285,85	-1,62	0,000492	1
226120_at	TTC8: tetrapeptide repeat domain 8	AW293939	1193,74	443,22	-2,69	1,91E-06	0,10434302
226207_at	FLJ39378: hypothetical protein FLJ39378	AI358954	792,25	392,52	-2,02	9,11E-05	1
226518_at	KCTD10: potassium channel tetramerisation domain containing 10	AW073741	1248,05	701,15	-1,78	0,000227	1
226546_at	CDNA clone IMAGE:5268696	BG477064	787,65	274,6	-2,87	1,02E-05	0,5547459
226565_at	TMEM99: transmembrane protein 99	AW054855	685,82	342,61	-2	0,000396	1
226790_at	MORN2: MORN repeat containing 2	AW015683	709,7	364,72	-1,95	7,67E-06	0,41817403
226833_at	CYB5D1: cytochrome b5 domain containing 1	AI921877	434,78	192,52	-2,26	0,000381	1
226938_at	WDR21A: WD repeat domain 21A	AA160604	373,21	194,62	-1,92	0,000108	1
226966_at	PRPF40B: PRP40 pre-mRNA processing factor 40 homolog B (S. cerevisiae)	BF108696	224,76	111,91	-2,01	0,00014	1
226970_at	FBXO33: F-box protein 33	AI690694	592,86	270,44	-2,19	5,78E-05	1
227033_at	PDI3: protein disulfide isomerase family A, member 3	AI825800	636,52	310,11	-2,05	0,000667	1
227062_at	TncRNA: trophoblast-derived noncoding RNA	AU155361	583,22	4861,63	8,34	0,000627	1
227070_at	GLT8D2: glycosyltransferase 8 domain containing 2	W63754	2677,02	371,68	-7,2	1,35E-05	0,73400534
227091_at	KIAA1505: KIAA1505 protein	AB040938	443,89	170,12	-2,61	0,000554	1
227134_at	SYTL1: synaptotagmin-like 1	AI341537	77,93	384,57	4,93	0,000144	1
227153_at	IMMP2L: IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)	AI784580	689,79	316,98	-2,18	0,000148	1
227294_at	ZNF689: zinc finger protein 689	AI474448	234,17	129,6	-1,81	0,000732	1
227337_at	ANKRD37: ankyrin repeat domain 37	AA886870	455,98	171,57	-2,66	0,000654	1
227359_at	C1orf102: chromosome 1 open reading frame 102	AI911248	235,15	81,66	-2,88	0,000108	1
227431_at	CDNA clone IMAGE:4791585	BF435958	892,5	411,31	-2,17	0,000765	1
227521_at	FBXO33: F-box protein 33	N22902	380,84	155,79	-2,44	1,22E-05	0,66678828
227572_at	USP30: ubiquitin specific peptidase 30	AA528138	654,52	317,7	-2,06	3,93E-05	1
227580_s_at	DKFZP434B0335: DKFZP434B0335 protein	BE616972	457,17	285,14	-1,6	0,000751	1
227657_at	RNF150: ring finger protein 150	AA722069	188,96	47,66	-3,96	2,19E-05	1
227740_at	UHMK1: U2AF homology motif (UHM) kinase 1	AW173222	415,62	1012,55	2,44	0,00013	1
227752_at	SPTLC3: serine palmitoyltransferase, long chain base subunit 3	AA005105	196,07	66,18	-2,96	0,000645	1
227760_at	IGFBPL1: insulin-like growth factor binding protein-like 1	AL522781	795,89	229,43	-3,47	0,000745	1
227889_at	AYTL1: acyltransferase like 1	AI765437	3380,99	1172,16	-2,88	0,000314	1
227983_at	MGC7036: hypothetical protein MGC7036	AI810244	1328,64	724,74	-1,83	0,000433	1

228064_at	C22orf36: chromosome 22 open reading frame 36	AW006520	323,02	104,92	-3,08	0,000161	1
228249_at	C11orf74: chromosome 11 open reading frame 74	AA535128	1030,31	313,17	-3,29	5,35E-05	1
228407_at	SCUBE3: signal peptide, CUB domain, EGF-like 3	AI733234	4433,16	1486,23	-2,98	0,000437	1
228429_x_at	KIF9: kinesin family member 9	BG168764	454,34	224,66	-2,02	0,000139	1
228811_at	Transcribed locus	AI493276	234,62	83,27	-2,82	0,000266	1
229024_at	CDNA FLJ10151 fis, clone HEMBA1003402	BF056892	314,64	75,72	-4,16	8,79E-06	0,47932082
229032_at	WSCD2: WSC domain containing 2	BE962770	665,52	182,44	-3,65	6,43E-05	1
229238_at	LOC400566: hypothetical gene supported by AK128660	BE552331	399,9	165,31	-2,42	0,000395	1
229272_at	FBNP4: formin binding protein 4	AI083506	62,29	205,96	3,31	0,000473	1
229287_at	PCNX: pecanex homolog (Drosophila)	BE326214	391,78	218,84	-1,79	0,000532	1
229415_at	CYCS: cytochrome c, somatic	BF593856	255,25	97,45	-2,62	0,000309	1
229452_at	TMEM88: transmembrane protein 88	AL544576	213,94	74,43	-2,87	5,45E-06	0,29757551
229686_at	P2RY8: purinergic receptor P2Y, G-protein coupled, 8	AI436587	357,92	152,9	-2,34	4,56E-05	1
229986_at	ZNF717: zinc finger protein 717	AW205616	230,99	118,86	-1,94	0,000248	1
230031_at	HSPA5: heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	AW052044	1696,74	694,97	-2,44	0,000838	1
230290_at	SCUBE3: signal peptide, CUB domain, EGF-like 3	BE674338	582,22	214,44	-2,71	0,000499	1
230387_at	Transcribed locus	AL038450	111,26	340,7	3,06	0,000304	1
230552_at	LOC284412: hypothetical protein LOC284412	AI936524	246,31	105,21	-2,34	0,00057	1
230651_at	Transcribed locus	AI018256	55,86	218,33	3,91	0,000309	1
230669_at	RASA2: RAS p21 protein activator 2	W38444	140,86	326,79	2,32	4E-05	1
230976_at	C9orf98: chromosome 9 open reading frame 98	AW663881	174,88	69,64	-2,51	1,72E-06	0,09384959
231166_at	GPR155: G protein-coupled receptor 155	AI733474	343,93	79,81	-4,31	9,57E-06	0,52202197
	Transcribed locus, weakly similar to XP_001107312.1						
	dihydrolipoamide branched chain transacylase isoform 3						
231199_at	[Macaca mulatta]	AA701676	48,43	185,49	3,83	0,00015	1
231240_at	DIO2: deiodinase, iodothyronine, type II	AI038059	7841,43	1971,61	-3,98	0,000559	1
231358_at	Transcribed locus	BE465760	245,85	75,71	-3,25	7,95E-05	1
231431_s_at	MRNA; cDNA DKFZp762E1314 (from clone DKFZp762E1314)	AI125670	397,98	207,24	-1,92	0,000834	1
231530_s_at	C11orf1: chromosome 11 open reading frame 1	BG150085	755,25	354,26	-2,13	0,00019	1
231838_at	C20orf119: chromosome 20 open reading frame 119	AK026760	48,38	160,65	3,32	0,000518	1
231870_s_at	NMD3: NMD3 homolog (S. cerevisiae)	BG291007	615,68	1093,68	1,78	0,000106	1
232001_at	LOC439949: hypothetical gene supported by AY007155	AW193600	205,17	84,1	-2,44	8,95E-05	1

232053_x_at	RHBD2: rhomboid domain containing 2	AL533352	1522,93	684,56	-2,22	0,000512	1
232146_at	NDUFC1: NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	AK023115	248,81	79,54	-3,13	7,28E-06	0,39728962
232150_at	CDNA clone IMAGE:4792085	AA134418	69,8	194,35	2,78	0,000101	1
232264_at	CDNA FLJ12142 fis, clone MAMMA1000356	AK022204	82,14	330,36	4,02	0,000725	1
232489_at	CCDC76: coiled-coil domain containing 76	AK001149	63,27	189,7	3	5,23E-05	1
232541_at	CDNA FLJ20099 fis, clone COL04544	AK000106	137,43	712,65	5,19	0,000147	1
233595_at	USP34: ubiquitin specific peptidase 34	AK024341	269,94	639,22	2,37	0,000362	1
234302_s_at	ALKBH5: alkB, alkylation repair homolog 5 (E. coli)	AL137263	1765,96	1064,59	-1,66	0,00046	1
234972_at	ARL16: ADP-ribosylation factor-like 16	BE746724	446,3	269,4	-1,66	0,00047	1
235003_at	UHMK1: U2AF homology motif (UHM) kinase 1	AI249980	144,22	288,23	2	0,000289	1
235010_at	LOC729013: hypothetical protein LOC729013	AA833832	685,24	313,59	-2,19	0,000652	1
235115_at	PDE8B: phosphodiesterase 8B	AV722254	182,56	59,68	-3,06	0,000152	1
235136_at	ORMDL3: ORM1-like 3 (S. cerevisiae)	BF337528	309,77	118,56	-2,61	0,000331	1
235264_at	Transcribed locus	AW956392	441,39	218,15	-2,02	0,000446	1
235282_at	CDNA clone IMAGE:5206119	BF447113	331,91	138,53	-2,4	8,41E-05	1
235391_at	FAM92A1: family with sequence similarity 92, member A1	AW960748	62,85	163,8	2,61	0,000241	1
235757_at	Transcribed locus	AA814006	115,15	275,9	2,4	0,000705	1
236035_at	Transcribed locus	AW190406	201,26	63,66	-3,16	0,000337	1
236038_at	Transcribed locus	N50714	226,42	74,11	-3,06	8,85E-05	1
236115_at	Transcribed locus, moderately similar to XP_517655.1 similar to KIAA0825 protein [Pan troglodytes]	AA035771	162,31	47,66	-3,41	7,29E-06	0,39746017
236201_at	Transcribed locus	N30188	348,63	103,37	-3,37	0,000523	1
236359_at	SCN4B: sodium channel, voltage-gated, type IV, beta	AW026241	240,35	84,73	-2,84	0,000349	1
236472_at	Transcribed locus	AI806586	74,65	204,82	2,74	8,32E-05	1
236696_at	SR140: U2-associated SR140 protein	BE464843	41,64	142,42	3,42	0,000403	1
236741_at	WDR72: WD repeat domain 72	AW299463	197,25	73,55	-2,68	0,000431	1
236841_at	FAM39DP: Family with sequence similarity 39, member D pseudogene	BE464132	190,69	469,96	2,46	8,92E-05	1
237421_at	Full length insert cDNA clone ZD48A05	BF509605	229,79	24,28	-9,46	0,000566	1
237741_at	SLC25A36: Solute carrier family 25, member 36	AW514168	96,9	212,01	2,19	0,000374	1
238081_at	C4orf12: chromosome 4 open reading frame 12	AI694300	216,02	76,2	-2,83	0,000683	1
238156_at	Transcribed locus	AW205632	88,01	205,8	2,34	0,000749	1

238563_at	Transcribed locus	AV762916	61,19	197,31	3,22	8,61E-05	1
238761_at	ELK4: ELK4, ETS-domain protein (SRF accessory protein 1)	BE645241	327,2	628,46	1,92	0,000243	1
239533_at	GPR155: G protein-coupled receptor 155	AI970061	199,48	73,8	-2,7	0,000436	1
239587_at	Transcribed locus	AI686890	166,56	61,26	-2,72	0,000624	1
239598_s_at	AYTL1: acyltransferase like 1	AA789296	485,14	164,12	-2,96	5,55E-05	1
239763_at	PRDM11: PR domain containing 11	AA157112	187,95	68,97	-2,72	5,02E-05	1
239893_at	Transcribed locus	AA702409	41,11	142,5	3,47	0,000384	1
240065_at	FAM81B: family with sequence similarity 81, member B	AI769413	226,21	19,69	-11,49	1,18E-05	0,64161256
240137_at	Transcribed locus	AI915629	192,49	38,92	-4,95	2,05E-05	1
240236_at	Transcribed locus	N50117	128,71	24,19	-5,32	0,000155	1
240307_at	Transcribed locus	N54783	58,91	166,66	2,83	0,000817	1
240395_at	CDNA FLJ42406 fis, clone ASTRO3000482	AI635761	850,44	180,49	-4,71	0,000384	1
240594_at	Transcribed locus	W86659	62,97	211,49	3,36	0,00018	1
241401_at	C4orf12: chromosome 4 open reading frame 12	BG496631	173,37	52,95	-3,27	7,28E-05	1
241681_at	Transcribed locus	AW296451	105,54	351,77	3,33	0,000738	1
241885_at	Transcribed locus	BF431050	29,45	141,8	4,82	1,24E-05	0,67570339
242057_at	Transcribed locus	AI301859	205	58,83	-3,48	1,44E-05	0,78295417
242068_at	Transcribed locus	AA608834	78,37	214,23	2,73	0,000372	1
242133_s_at	LOC654342: Similar to lymphocyte-specific protein 1	AA630955	65,78	173,48	2,64	0,000311	1
242146_at	SNRPA1: Small nuclear ribonucleoprotein polypeptide A'	AA872471	87,42	243,31	2,78	0,000332	1
242191_at	NBPF10 /// NBPF11: neuroblastoma breakpoint family, member 11 /// neuroblastoma breakpoint family, member 10	AI701905	186,17	467,2	2,51	0,000822	1
242233_at	Transcribed locus	AI739332	93,08	213,14	2,29	0,000604	1
242282_at	ZFPM1: zinc finger protein, multitype 1	AI889717	412,91	200,35	-2,06	0,00033	1
242471_at	Clone HLS_IMAGE_238756 mRNA sequence	AI916641	51,26	206,35	4,03	0,000598	1
242903_at	IFNGR1: interferon gamma receptor 1	AI458949	64,3	218,64	3,4	0,000642	1
243110_x_at	NPW: neuropeptide W	AI868441	37,86	235,9	6,23	4,61E-05	1
243304_at	LOC286109: hypothetical protein LOC286109	AI733824	38,1	165,64	4,35	0,000152	1
243909_x_at	GUSB2: glucuronidase, beta-like 2	R43205	106,73	223,04	2,09	0,000779	1
243917_at	CLIC5: chloride intracellular channel 5	AW083491	394,2	680,51	1,73	0,000212	1
244341_at	Transcribed locus	AA827728	59,12	233,18	3,94	0,000159	1
244587_at	Hs.105820.0	AA534039	261,06	141,85	-1,84	0,000171	1
244803_at	Transcribed locus	AI335191	66,23	178,81	2,7	0,000586	1

244826_at	Transcribed locus	R24061	93,56	238,77	2,55	0,000779	1
43977_at	TMEM161A: transmembrane protein 161A	AI660497	358,44	202,13	-1,77	0,000738	1
51146_at	PIGV: phosphatidylinositol glycan anchor biosynthesis, class V	AA203365	737,85	361,28	-2,04	4,67E-06	0,25507399
59437_at	C9orf116: chromosome 9 open reading frame 116	AI830563	182,86	50,06	-3,65	2,82E-06	0,15415449
61297_at	CASKIN2: CASK interacting protein 2	AL037338	222,45	116,78	-1,9	0,000713	1
64900_at	FLJ22167: hypothetical protein FLJ22167	AA401703	447,06	183,97	-2,43	0,000384	1

Claims

1. An mRNA classifier for characterising a sample obtained from a thyroid nodule of an individual, wherein said mRNA classifier
 - 5 i) consists of the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1,

10 LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330 and NA; or

15 ii) comprises at least all of the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1,

20 LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330 and NA; or

25 iii) comprises six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR,

30 BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM,

- CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, SLCO2A1 and NA; wherein at least six mRNAs selected from the subgroup consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF are included; or
- 5 iv) comprises or consists of six or more mRNAs selected from the group consisting of the mRNAs disclosed in table 19;
- and distinguishes between the classes thyroid follicular adenoma and thyroid follicular carcinoma, wherein said distinction is given as a prediction probability for said sample of belonging to either class, said probability being a number falling in the range of from 0 to 1.
- 10 2. The mRNA classifier according to claim 1, wherein said mRNA classifier comprises or consists of between 6 to 10 mRNAs, such as 5 to 10, for example 10 to 15, such as 15 to 20, for example 20 to 25, such as 25 to 30, for example 30 to 35, such as 35 to 40, for example 40 to 45, such as 45 to 50, for example 50 to 55, such as 55 to 60, for example 60 to 65, such as 65 to 70, for example 70 to 75, such as 75 to 80, for example 80 to 85, such as 85 to 90, for example 90 to 95, such as 95 to 100, for example 100 to 110, such as 110 to 120 mRNAs.
- 15 3. The mRNA classifier according to any of the preceding items, wherein the sensitivity is at least 85%, such as at least 86%, for example at least 87%, such as at least 88%, for example at least 89%, such as at least 90%, for example at least 91%, such as at least 92%, for example at least 93%, such as at least 94%, for example at least 95%.
- 20 4. The mRNA classifier according to any of the preceding items, wherein the specificity is at least 85%, such as at least 86%, for example at least 87%, such as at least 88%, for example at least 89%, such as at least 90%, for example at least 91%, such as at least 92%, for example at least 93%, such as at least 94%, for example at least 95%.
- 30

5. The mRNA classifier according to any of the preceding items, wherein the prediction probability of a sample for belonging to a certain class is a number falling in the range of from 0 to 1, such as from 0.0 to 0.1, for example 0.1 to 0.2, such as 0.2 to 0.3, for example 0.3 to 0.4, such as 0.4 to 0.49, for example 0.5, such as 0.51 to 0.6, for example 0.6 to 0.7, such as 0.7 to 0.8, for example 0.8 to 0.9, such as 0.9 to 1.0.
5
6. The mRNA classifier according to any of the preceding items, wherein an alteration of the expression profile of one or more of said mRNAs is associated with thyroid follicular carcinoma, or thyroid follicular adenoma, or fetal adenoma, or thyroid follicular carcinoma and fetal adenoma.
10
7. The mRNA classifier according to any of the preceding items, wherein the expression level of one or more mRNAs is determined by the microarray technique.
15
8. The mRNA classifier according to any of the preceding items, wherein the expression level of one or more mRNAs is determined by the quantitative polymerase chain reaction (QPCR) technique.
20
9. The mRNA classifier according to any of the preceding items, wherein the sample is extracted from an individual by fine-needle aspiration.
10. A model for predicting the diagnosis of an individual with a thyroid nodule, comprising
25
 - i) providing a set of input data to the mRNA classifier according to any of the preceding items, and
 - ii) determining if said individual has a condition selected from the group of thyroid follicular adenoma, thyroid follicular carcinoma, and/or fetal adenoma.
30
11. A device for measuring the expression level of mRNAs in a sample from a thyroid nodule, wherein said device
 - i) consists of probes or probe sets for mRNAs selected from the groups consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1,
35

- LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330 and NA; or
- 5
- 10 ii) comprises probes or probe sets for at least all of the mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330 and NA; or
- 15
- 20
- iii) comprises probes or probe sets for at least six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, SLCO2A1 and NA; wherein at least six mRNAs selected from
- 25
- 30
- 35

the subgroup consisting of TOP2A, RRM2, PBK, ANLN, NR4A1, FOSB, EGR2 and CTGF are included; or

- iv) comprises or consists of six or more mRNAs selected from the group consisting of the mRNAs disclosed in table 19;

5 wherein said device is used for classifying a sample obtained from a thyroid nodule of an individual.

10 12. The device according to claim 11, wherein said device is used with the mRNA classifier according to any of claims 1 to 9 to classify a sample into either of the classes of thyroid follicular adenoma, or thyroid follicular carcinoma, or fetal adenoma, or thyroid follicular adenoma merged with fetal adenoma.

15 13. The device according to claim 11, wherein said device comprises or consists of between 6 to 10 probes or probe sets for mRNAs, such as 5 to 10, for example 10 to 15, such as 15 to 20, for example 20 to 25, such as 25 to 30, for example 30 to 35, such as 35 to 40, for example 40 to 45, such as 45 to 50, for example 50 to 55, such as 55 to 60, for example 60 to 65, such as 65 to 70, for example 70 to 75, such as 75 to 80, for example 80 to 20 85, such as 85 to 90, for example 90 to 95, such as 95 to 100, for example 100 to 110, such as 110 to 120 probes or probe sets.

25 14. The device according to any of claims 11 to 13, wherein said device is a microarray chip.

15. The device according to any of claims 11 to 13, wherein said device is a QPCR Microfluidic Card, QPCR tubes, QPCR tubes in a strip or a QPCR plate.

30 16. A kit-of-parts comprising the device of any of claims 11 to 15, and at least one additional component.

35 17. The kit according to claim 16, wherein said additional component is means for extracting RNA, such as mRNA, from a sample.

18. The kit according to claim 16, wherein said additional component is reagents for performing microarray analysis and/or reagents for performing QPCR analysis.

- 5 19. A method for diagnosing if an individual has, or is at risk of developing, follicular thyroid carcinoma and/or fetal adenoma, comprising the steps of:
- i) extracting RNA from a sample obtained from a thyroid nodule of an individual,
 - 10 ii) analysing the mRNA expression profile of the sample, wherein said mRNAs
 - i. consists of the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2,
 15 BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330 and NA; or
 20
 - ii. comprises at least all of the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2,
 25 TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330 and NA; or
 30
 - iii. comprises or consists of six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20,
 35 CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD,

5 CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF,
 NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1,
 LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80,
 PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY,
 APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91,
 C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1,
 SH3RF1, EBAG9, MCFD2, PLDN, TCEAL4, ZNF330, ASPM,
 CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1,
 10 SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1,
 DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4,
 SLCO2A1 and NA; or

iv. comprises or consists of six or more mRNAs selected from the
 group consisting of the group disclosed in table 19;
 wherein a predetermined mRNA expression profile of said mRNAs is
 15 indicative of the individual having, or being at risk of developing, follicular
 thyroid carcinoma and/or fetal adenoma.

20. The method according to claim 19, wherein said method comprises
 obtaining prediction probabilities of between 0-1 for said sample.

20

21. A system for performing a diagnosis on an individual with a thyroid nodule,
 comprising:

i) means for analysing the mRNA expression profile of the thyroid
 nodule, wherein said mRNAs

25 i. consists of the group consisting of FOSB, LOC286002, CA4,
 EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4,
 MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5,
 NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN,
 EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK,
 30 HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2,
 CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL,
 ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11,
 KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B,
 TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCFD2,
 35 PLDN, TCEAL4, ZNF330 and NA; or

- ii. comprises at least all of the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCDF2, PLDN, TCEAL4, ZNF330 and NA; or
 - iii. comprises or consists of six or more mRNAs selected from the group consisting of FOSB, LOC286002, CA4, EGR2, PLA2R1, LMOD1, DNASE1L3, PTPRN2, ZMAT4, MAN1C1, ARHGAP20, CTGF, SDPR, CCDC85A, ITIH5, NR4A1, MPPED2, HGD, CITED2, RRM2, TOP2A, ANLN, EZH2, BIRC5, CENPF, NUSAP1, UBE2C, CCNB2, MELK, HMMR, BUB1B, BUB1, LOC100131139, LMNB1, HIG2, CDCA3, XPR1, KRT80, PAFAH1B3, RCC2, CTDSPL, ARPC5L, CBX3, H2AFY, APOLD1, C13orf15, COLEC11, KIAA1467, MAFB, C17orf91, C4orf12, SPARCL1, MYO15B, TMEM88, IVD, CENTD1, AAK1, SH3RF1, EBAG9, MCDF2, PLDN, TCEAL4, ZNF330, ASPM, CDCA5, CEP55, CKS2, CTD, H2A, KIF4A, NEK2, PBK, PRC1, SAC3D1, TMPO, TPX2, AGTR1, CDH16, CYR61, DLC1, DUSP14, FOSB, JUN, KCNAB1, MATN2, NR4A3, SLC26A4, SLCO2A1 and NA; or
 - iv. comprises or consists of six or more mRNAs selected from the group consisting of the group disclosed in table 19; and
 - ii) means for determining if said individual has a benign or malignant and/or pre-malignant condition selected from follicular thyroid adenoma, follicular thyroid carcinoma and fetal adenoma.

22. A computer program product having a computer readable medium, said computer program product providing a system for predicting the diagnosis of an individual with a thyroid nodule, said computer program product

comprising means for carrying out any of the steps of the system according to claim 21.

- 5 23. A system according to claim 21, wherein the data is stored, such as stored in at least one database.

1/9

Figure 1

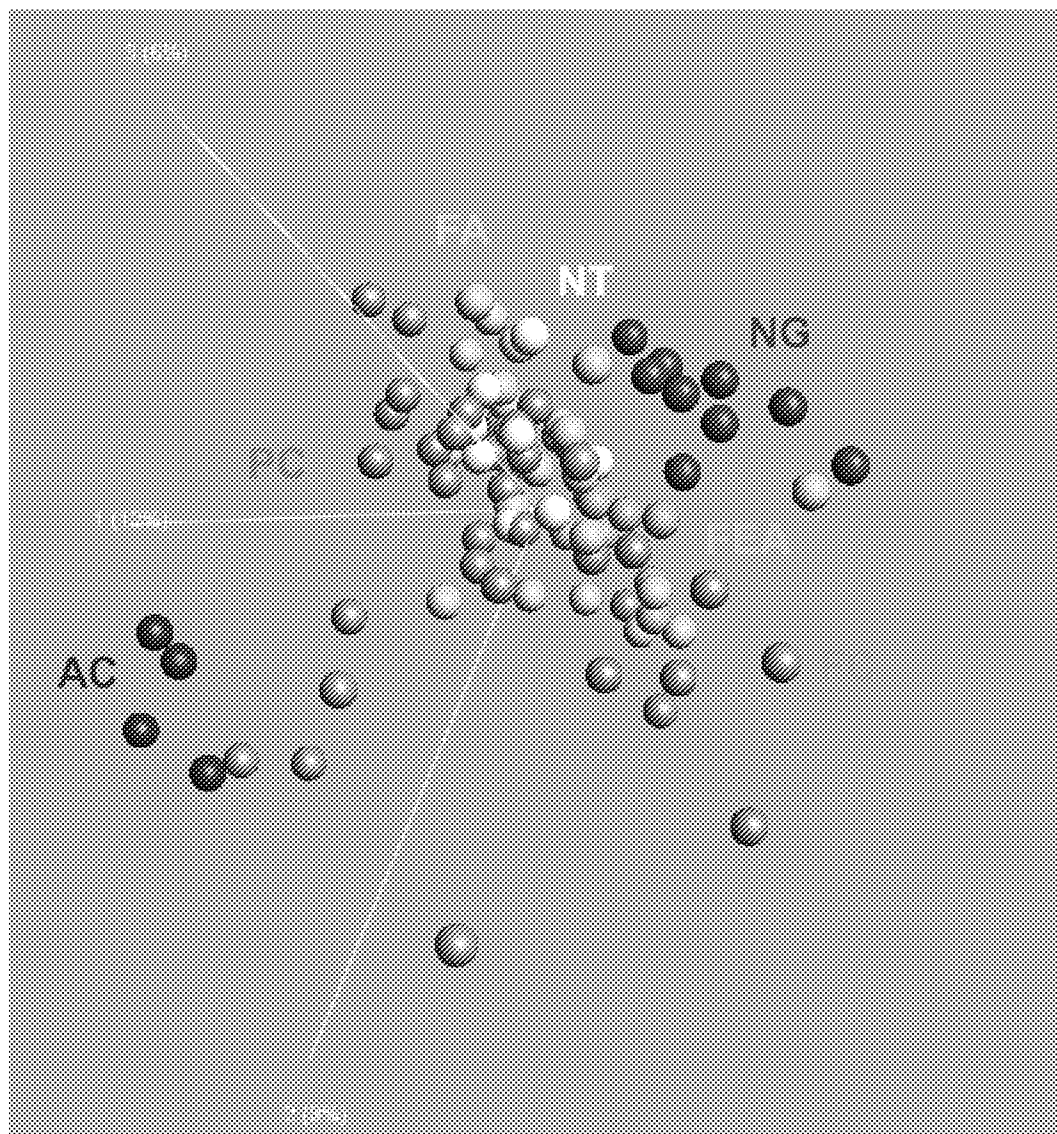
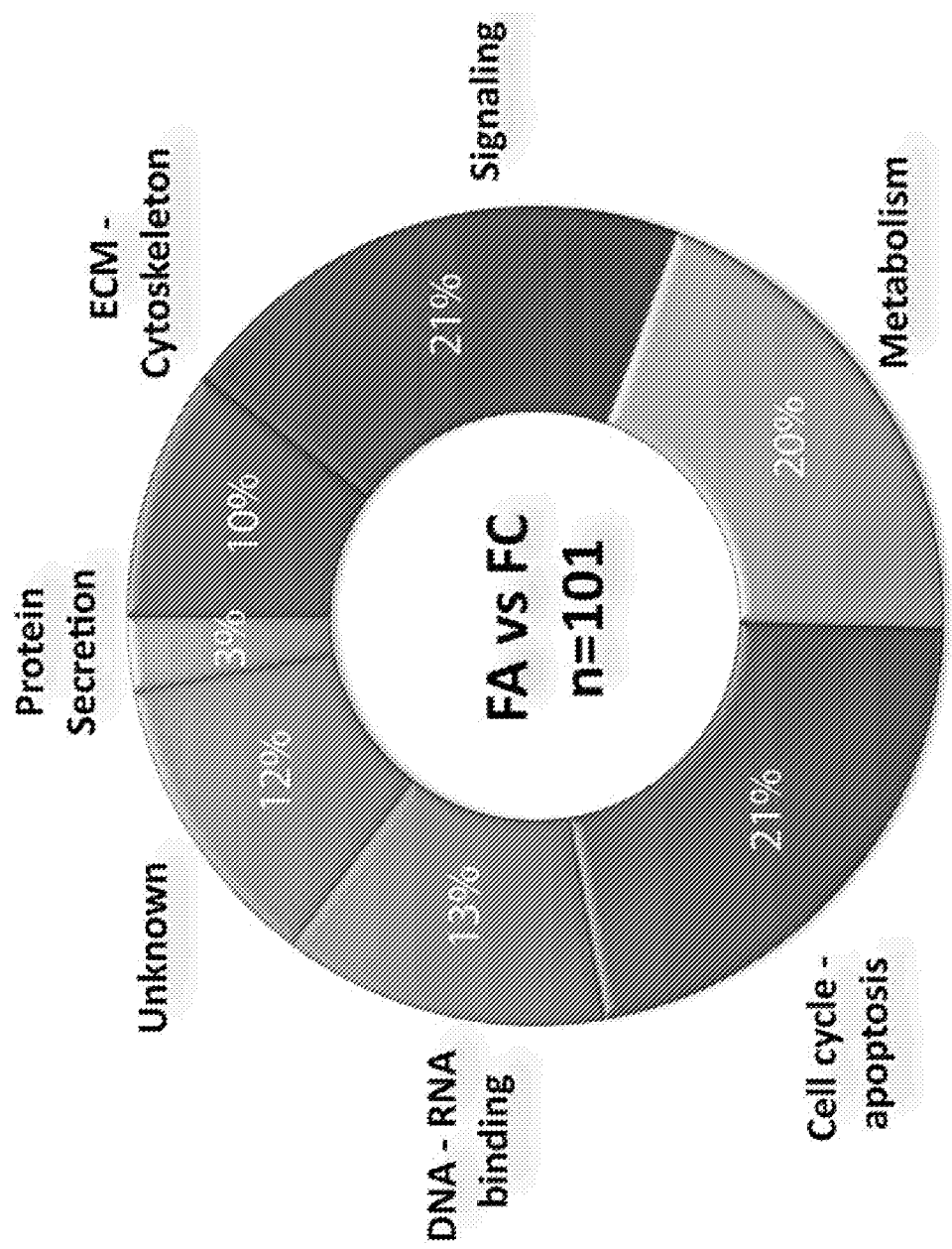


Figure 2



3/9

Figure 3

Cell cycle and mitosis	FC	P-value
ANLN: anillin	26	0.013
ARP3L: actin related protein 3/2 complex, subunit 5-like	2	0.023
ASPM: asp (abnormal spindle) homolog	26	0.034
BUB1B: BUB1 budding uninhibited by benzimidazoles 1	9	0.029
CBX3: chromobox homolog 3	2	0.020
CCNB2: cyclin B2	14	0.017
CDC45: cell division cycle associated 5	12	0.048
CENPF: centromere protein F	18	0.027
CEP55: centrosomal protein 55kDa	15	0.038
CKS2: CDC28 protein kinase regulatory subunit 2	4	0.038
CTD: carboxy-terminal domain, RNA polymerase II	2	0.021
H2A histone family, member Y	2	0.017
KIF4A: kinesin family member 4A	18	0.037
MEIK: maternal embryonic leucine zipper kinase	16	0.026
NEK2: NIMA (never in mitosis gene a)-related kinase 2	15	0.031
NUSAP1: nucleolar and spindle associated protein 1	10	0.020
POK: POZ binding kinase	36	0.038
PRC1: protein regulator of cytokinesis 1	10	0.043
RCC2: regulator of chromosome condensation 2	2	0.020
RRM2: ribonucleotide reductase M2 polypeptide	20	0.005
SAC3D1: SAC3 domain containing 1	3	0.043
TMPO: thymopoietin	4	0.046
TOP2A: topoisomerase (DNA) II alpha 170kDa	18	0.002
TPX2: microtubule-associated, homolog	23	0.021
UBE2C: ubiquitin-conjugating enzyme E2C	15	0.030
Apoptosis and growth arrest	FC	P-value
AGTR1: angiotensin II receptor	-7	0.046
CCDC85A: coiled-coil domain containing 85A	-2	0.018
CDH16: cadherin 16	-4	0.041
CITED2: Cbp/p300-interacting transactivator	-3	0.008
CTGF: connective tissue growth factor	-6	0.011
CYR61: cysteine-rich, angiogenic inducer, 61	-4	0.038
DLC1: deleted in liver cancer 1	-3	0.042
DNASE1L3: deoxyribonuclease I-like 3	-7	0.005
DUSP14: dual specificity phosphatase 14	-2	0.035
EGR2: early growth response 2	-8	0.005
FOSB: FBJ murine osteosarcoma viral oncogene	-8	0.010
JUN: jun oncogene	-2	0.034
KCNAB1: potassium voltage-gated channel	-4	0.045
MAN1C1: mannosidase, alpha	-3	0.026
MATN2: matrilin 2	-3	0.043
NR4A1: nuclear receptor subfamily 4	-7	0.009
NR4A3: nuclear receptor subfamily 4, group A	-24	0.032
PLA2R1: phospholipase A2 receptor 1	-4	0.007
PTPRN2: protein tyrosine phosphatase	-2	0.027
SQPR: serum deprivation response	-5	0.010
SLC26A4: solute carrier family 26	-3	0.038
SLC22A1: solute carrier organic anion transporter family	-2	0.022

Figure 4

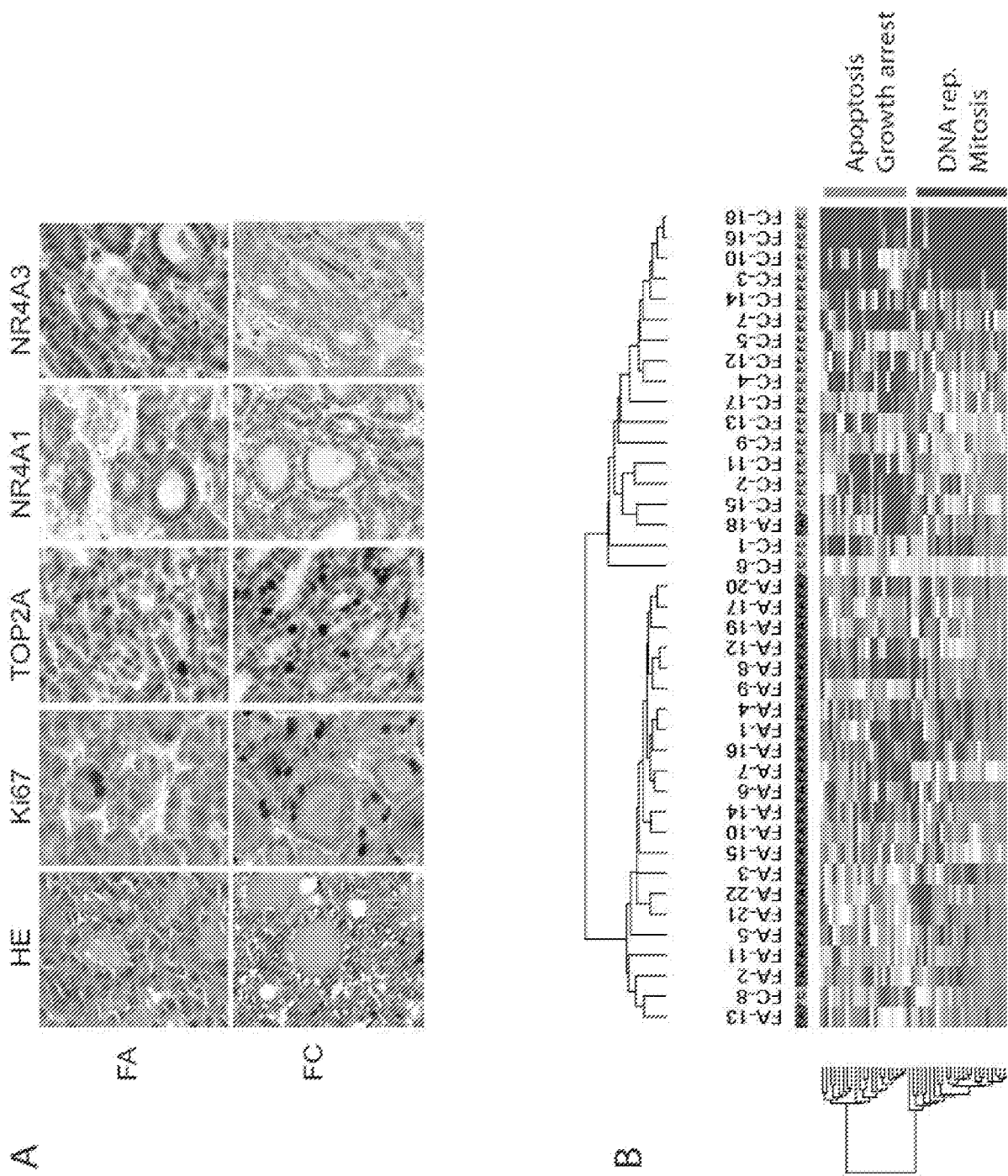
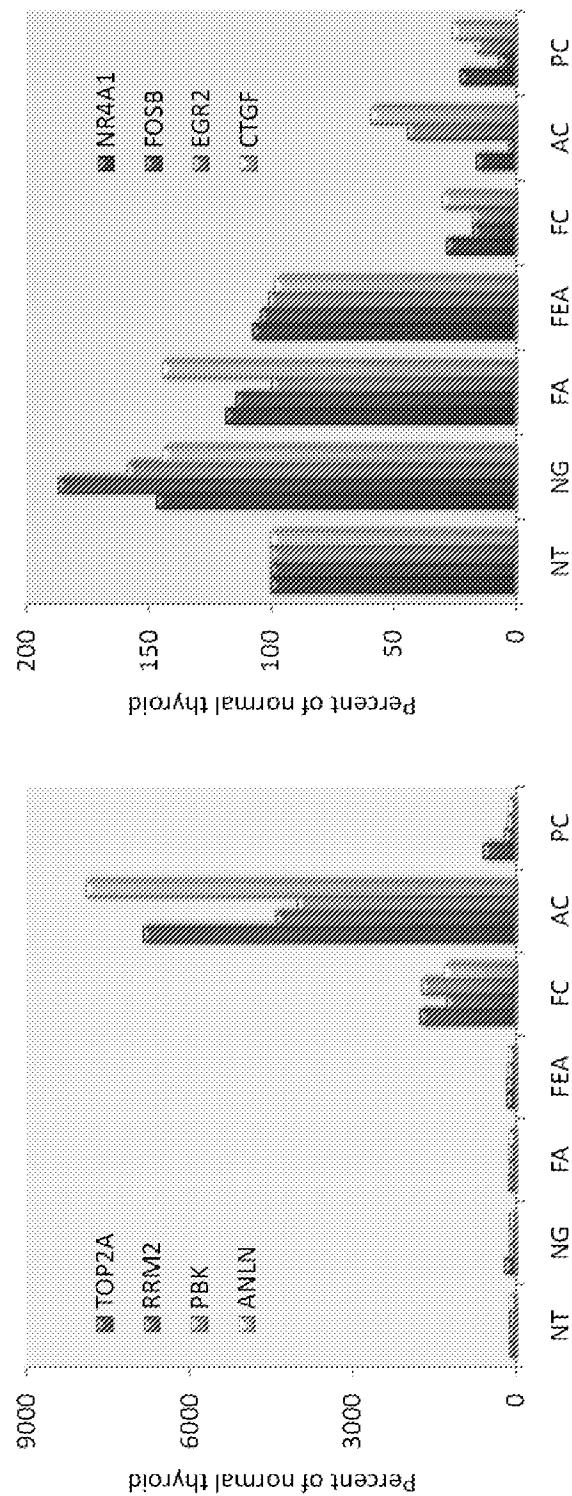
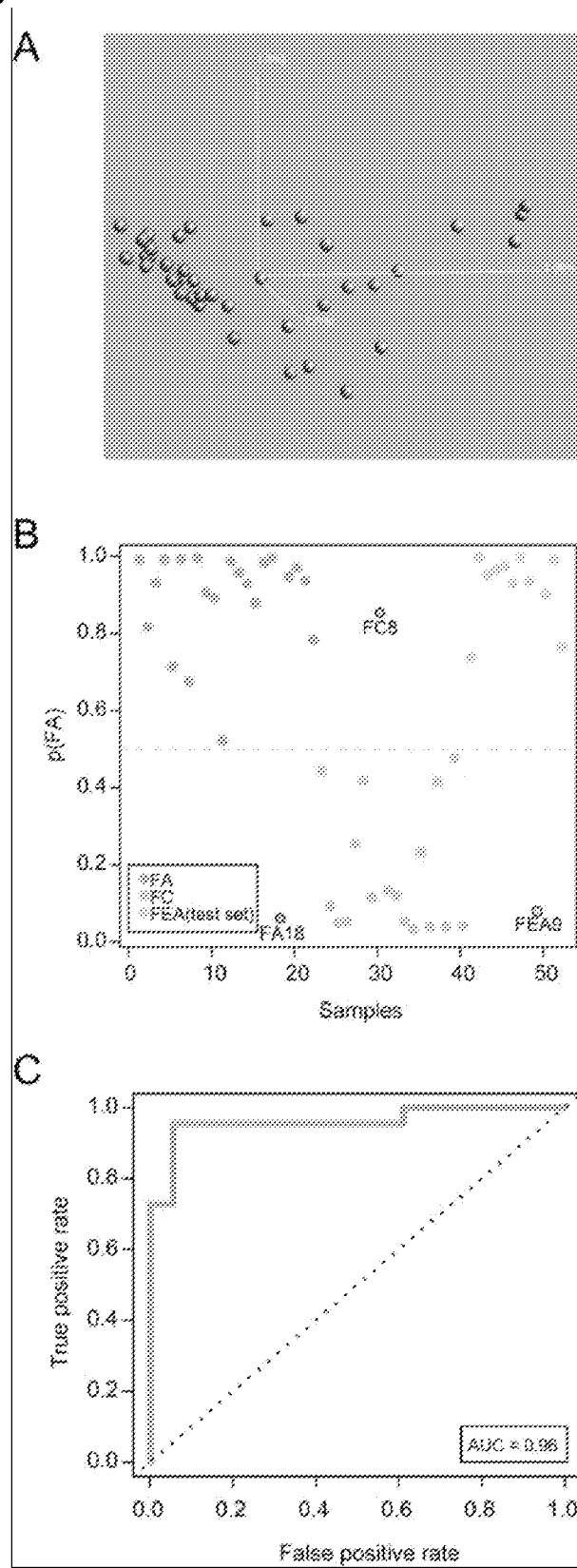


Figure 5



6/9

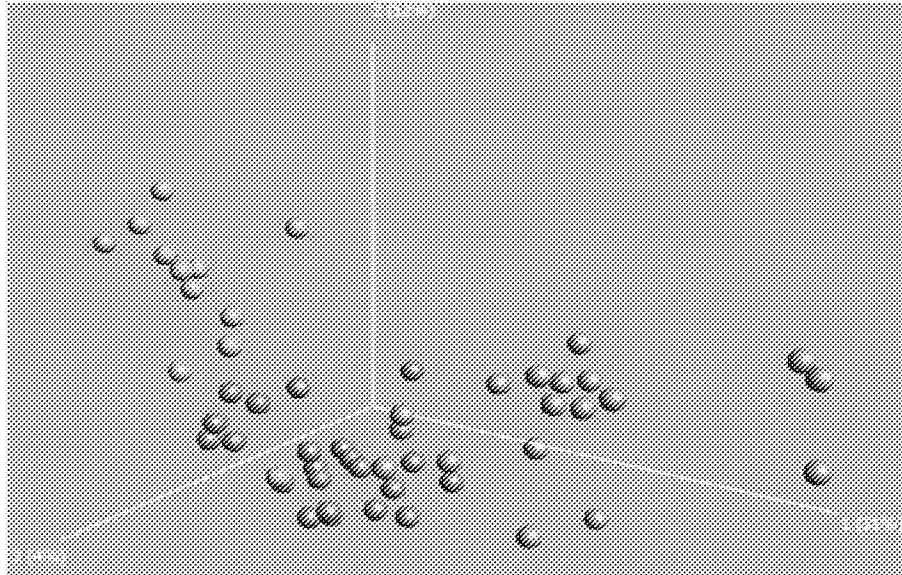
Figure 6



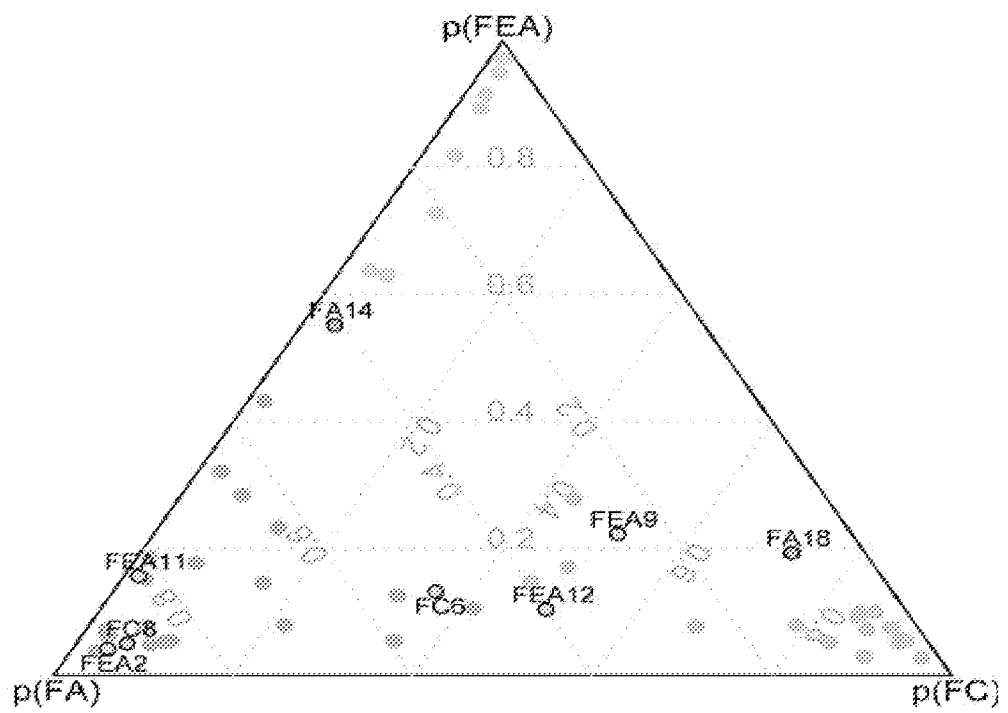
7/9

Figure 7

A

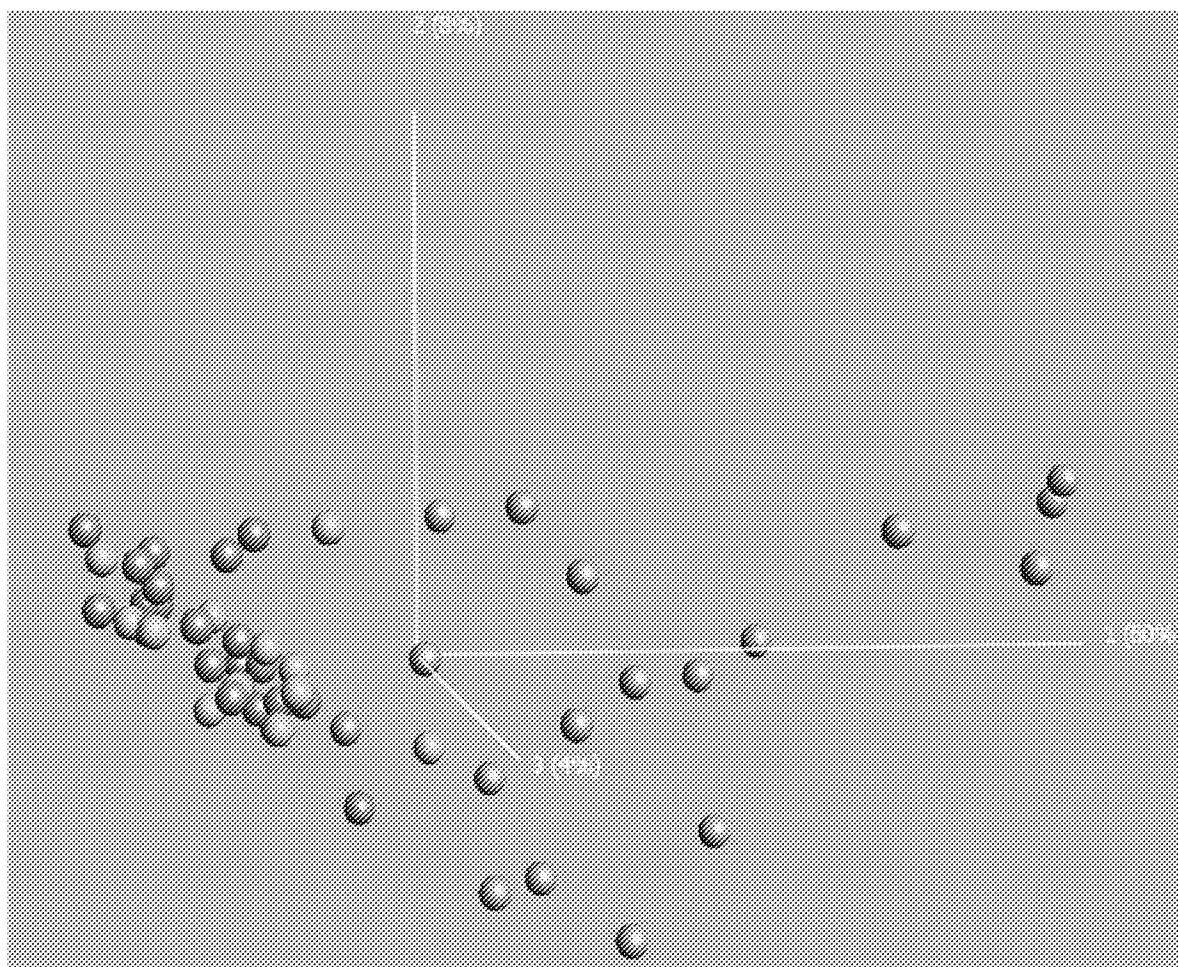


B



8/9

Figure 8



9/9

Figure 9

