Published in final edited form as:

Proteomics Clin Appl. 2008 October 10; 2(12): 1575–1585. doi:10.1002/prca.200780095.

Multiplex analysis of cytokines as biomarkers that differentiate benign and malignant thyroid diseases

Faina Linkov¹, Robert L. Ferris^{1,2,3,*}, Zoya Yurkovetsky¹, Adele Marrangoni¹, Lyudmila Velikokhatnaya¹, William Gooding¹, Brian Nolan¹, Matthew Winans¹, Eric R. Siegel⁴, Anna Lokshin¹, and Brendan C. Stack Jr⁵

¹Department of Medicine, University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA

²Department of Otolaryngology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

³Department of Immunology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

⁴Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

⁵Division of Head and Neck Oncology, Department of Otolaryngology-Head and Neck Surgery, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Abstract

Thyroid cancer incidence is increasing, and its diagnosis can be challenging. Fine needle biopsy, the principal clinical tool to make a tissue diagnosis, leads to inconclusive diagnoses in up to 30% of the cases, leading to surgery. Advances in proteomics are improving abilities to diagnose malignant conditions using small samples of tissue or body fluids. We hypothesized that analysis of serum growth factors would uncover diagnostically informative differences between benign and malignant thyroid conditions. Using xMAP profiling, we evaluated concentrations of 19 cytokines, chemokines, and growth factors. We used sera from 23 patients with cancer (Malignant group), 24 patients with benign nodular thyroid disease (Benign group), and 23 healthy subjects (Normal group). In univariate analysis, five factors (epithelial growth factor, hepatocyte growth factor, Interleukins-5 and -8, and regulated upon activation, normally T-expressed and presumably secreted (RANTES) distinguished subjects with thyroid disease from the Normal group. In multivariate analysis, the set {Interleukin-8, hepatocyte growth factor, monocyte-induced γ interferon, interleukin-12 p40} achieved noteworthy discrimination between Benign and Malignant groups (area under the receiver operating characteristics curve was 0.81 (95% confidence interval: 0.65–0.90)). Multiplex panels of serum biomarkers may be promising tools to diagnose cancer in patients presenting with evidence of nodular thyroid disease.

Keywords

Cytokines; Hepatocyte growth factor; Interleukins; Monocyte-induced γ interferon; Thyroid cancer

*Additional corresponding author: Dr. Robert L. Ferris; E-mail: ferrisrl@upmc.edu. **Correspondence:** Dr. Brendan C. Stack, Jr, Department of Otolaryngology-Head and Neck Surgery, 4301 W. Markham Street, Slot 543, Little Rock, AR 72205, USA, **E-mail:** bstack@uams.edu, Fax: 11-501-686-8029.

The authors have declared no conflict of interest.

1 Introduction

Nodular thyroid disease, indicated by the presence of single or multiple nodules within the thyroid gland, remains a common clinical problem that increases with age, affecting 50% of 50 years old. Only 5% of these nodules are malignant. In 2007, there will be 33 550 new cases of thyroid cancer diagnosed in the United States, which will be 2.3% of all malignancies diagnosed that year [1]. During 2007, 1530 patients will die of thyroid cancer. Nearly two-thirds of people who are diagnosed with thyroid cancer are between the ages of 20 and 55. The incidence of thyroid cancer in the US is 0.01% and is increasing at 2% a year [2].

There are few risk factors that make humans more likely to develop thyroid cancer, which include: a diet low in iodine, ionizing radiation exposure, obesity, and hereditary factors [3–6]. Most people with thyroid cancer; however, have no apparent risk factors, and other people with one or more risk factors may never develop this disease (American Cancer Society (cited; available from:

http://www.cancer.org/docroot/CRI/content/CRI_2_4_2X_What_are_the_risk_factors_for_thyroid_cancer_43.asp). Fine needle aspiration biopsy (FNAB) commonly used as the initial diagnostic test for thyroid cancer, is safe, accurate, and cost effective [7,8]. The big disadvantage of FNAB is that approximately 30% of biopsies of thyroid nodules are indeterminate, insufficient, or suspicious for malignancy, requiring diagnostic thyroid surgery [9]. Although FNAB technology has reduced the number of diagnostic surgeries for thyroid neoplasms [10], there is still a large number of patients who undergo these invasive procedures [11].

FNAB cannot distinguish benign follicular lesions from carcinomas. Large-bore needle biopsy is no more accurate and is associated with more complications [12]. Many treatment centers suggest surgical excision of all indeterminate and follicular lesions to make a definitive histological diagnosis [13]. Fine needle aspirate specimens of insufficient quantity are often repeated [14]. Usually three unsuccessful attempts at biopsy will lead to surgical excision of the thyroid tissue. Costs and complications of some thyroid surgeries could be avoided if there were blood tests that had the power to properly characterize thyroid mass lesions as malignant *versus* benign.

Studies of some individual serologic factors mediating inflammatory processes, angiogenesis and tumor growth correlate with thyroid cancer development and progression. Benign and malignant thyroid disease often shows inflammatory dysregulation, which may be reflected in distinct serum cytokine profiles [15]. For example, IL-6 expression is related to aggressiveness of both papillary and medullary thyroid cancer [16]. Despite the growing number of reports indicating the complex interactions between inflammatory and angiogenic factors in the development of thyroid cancer, studies have concentrated on a small number of markers [17, 18]. Vesely *et al.* [19] examined serum levels of IGF-I, HGF, TGF β 1, bFGF, and VEGF in 28 patients with thyroid gland tumors (14 adenomas, 14 papillary carcinomas) and compared these concentrations with those in healthy people. The only current clinical marker of well-differentiated thyroid cancer, thyroglobulin, is exclusively valuable in the setting of total thyroidectomy and in antithyroglobulin antibody-negative patients [20]. Fajarewicz *et al* [21] used Affymetrix microarrays successfully to distinguish papillary thyroid carcinoma from benign thyroid nodules, but their method required thyroid tissue.

We hypothesized that a broad set of serum analytes would lend greater clinical utility to the differentiation of benign and malignant nodular thyroid disease. Thus, we used a novel multi-analyte xMAP profiling technology (Luminex Corporation, Austin, TX) that allows for simultaneous measurement of multiple biomarkers in serum of thyroid cancer patients and patients with benign or no thyroid disease [22]. The primary objective of this study was to characterize cytokine profiles in the systemic circulation of patients with benign and malignant

thyroid disease, and compare them to each other as a pilot exercise to study the feasibility of developing a biomarker panel that can distinguish effectively between the two groups. If used in conjunction with FNAB, such a serologic screening tool could provide valuable information in the event of an inadequate or nondiagnostic smear or reduce the need for surgical excision of indeterminate and follicular lesions in order to make definitive diagnoses.

2 Materials and methods

2.1 Patient population

The study population comprised patients treated for thyroid disease at the Division of Otolaryngology-Head and Neck Surgery, Pennsylvania State University/Milton S. Hershey Medical Center (Hershey, PA, USA). Since the anaplastic form of thyroid cancer is rare and has characteristics that are very different than other forms of thyroid cancer (fast-growing and poorly responsive to therapy), these cancers were not included in this study. Single specimens were obtained prior to treatment over an 11-month accrual period from 24 patients with active thyroid cancer and 24 patients with benign thyroid disease; one sample from a Malignant case was subsequently excluded for hemolysis. Data from 23 healthy subjects were also included, in order to provide a reference Normal group. The demographics and pathology reports of the study populations are shown in Tables 1A and 1B. Written informed consent was obtained from each subject, after approval by the Institutional Review Board.

2.2 Collection and storage of blood serum

Ten ml of peripheral blood was drawn from subjects using standardized phlebotomy procedures and allowed to clot. Handling and processing was similar for all groups of patients. Sera were separated by centrifugation within 2 h, and all specimens were immediately aliquotted, frozen, and stored in a -80°C freezer. No more than one freeze-thaw cycle was allowed for each sample.

2.3 Multiplex serum analysis

The Luminex xMAP™ serum assays were performed in 96-well microplate format. Analytes were chosen based on knowledge of pathophysiology of malignancy and markers reported in the literature. Multiplex bead-based immunoassays for cytokines were purchased from Invitrogen. Other assays for proteins were developed in the Luminex Core Facility of University of Pittsburgh Hillman Cancer Center (Pittsburgh, PA, USA) according to the protocol by Luminex Corporation. The analytes were: Eotaxin-1, interferon-alpha (IFN- α), interferon-gamma (IFN- γ), interleukin (IL) 10, 12 p40, 13, 15, 17, 1 α , 2, 4, 5, 6, 7, 8, interferon γ -inducible protein 10 (IP-10), monocyte chemotactic protein α (MCP-1), monocyte-induced γ interferon (MIG), macrophage inflammatory protein 1- α (MIP-1 α), macrophage inflammatory protein 1- β (MIP-1 β), regulated upon activation, normally T-expressed and presumably secreted (RANTES), tumor necrosis factor alpha (TNF- α), tumor necrosis factor receptor I (TNF-RI), tumor necrosis factor receptor II (TNF-RII), death receptor 5 (DR5), epithelial growth factor (EGF), basic fibroblast growth factor (bFGF), granulocyte colony stimulating factor (G-CSF), granulocyte monocyte colony stimulating factor (GM-CSF), hepatocyte growth factor (HGF), and vascular endothelial growth factor (VEGF).

Internal (spiked) control analytes were included for validation and standard curves. For intra-assay precision, CVs were calculated from six replicates of a control serum run on a single plate. For inter-assay precision, CVs were calculated from at least five separate experimental runs of the same control serum. The intra-assay variability was 3.5–7%. Inter-assay variability was 11–15%. Correlation with appropriate ELISA was 89–98% and recovery from serum was 80–110%. Table 2 provides a complete list of the 32 factors originally assayed *via* the xMAP procedure, 19 of which met the inclusion criteria for statistical analysis (see below).

2.4 Statistical analysis

To assess demographic imbalances, the reference Normal samples were compared to thyroid disease subjects (Benign and Malignant samples considered together), then Benign and Malignant samples were compared to each other. Fisher's exact test was used for sex imbalance, and the two-sided Wilcoxon rank sum (WRS) test was used for age imbalance. Serologic factors were analyzed if they met the following inclusion criteria (i) more than half the thyroid disease patient samples yielded numerical "within range" results and (ii) the "out-of-range" (OOR) results were unambiguously coded as OOR-low or OOR-high. For each factor, OOR-low results were set equal to 80% of the minimum within-range value, while OOR-high results were set equal to 125% of the maximum within-range value. Factor distributions were summarized in diagnosis group as medians and quartiles, and displayed as dot plots (Fig. 1). For univariate analysis of the factors, the Kruskal-Wallis (KW) test was used to compare Malignants, Benigns, and the reference Normals for any difference among groups, while the two-sided WRS test was used on thyroid disease subjects to compare Malignants-to-Benigns. Univariate tests were conducted at 1% α to adjust for the multiple comparisons without inflating Type II error in this modestly powered pilot study.

The reference Normals were excluded from multivariate analysis, in order to focus on the research objective of distinguishing Malignant from Benign cases among patients presenting with thyroid disease. Multivariate analysis of variance (MANOVA) was used on all nineteen factors to extract the discriminant vector, which points along the gradient of Benign *versus* Malignant separation in 19 dimensional "factor space". For this, factor concentrations were transformed to their $\log_2$ values, and further standardized to zero mean and unit variance to remove dynamic-range effects. After extraction, the discriminant vector was normalized to unit length and vector coefficients were squared to obtain the squared correlation (r^2) of each factor's alignment with the Benign *versus* Malignant gradient. Factors with the highest r^2 values were then used to create sets ("panels") of two, three, and four in number, and multivariate logistic regression was used to examine these factor panels for their comparative ability to classify samples as Benign or Malignant. The number of factors on a panel was limited to four or fewer to assure the validity of results by providing ten or more samples *per* factor [23,24]. To simulate each panel's large-scale diagnostic use, we averaged together the results of training and testing the classifier on a large number of training and test sets randomly generated with permutation resampling. The resampling algorithm consisted of the following steps for each factor panel examined: (i) randomly divide data into training and test sets, (ii) train the classifier on the training set, (iii) apply the trained classifier to the test set, (iv) store test-set prediction probabilities, and (v) repeat. 25% of samples (six Benign and five Malignant) were randomized into the test set during each iteration of the algorithm; 10 000 iterations ensured every sample would be randomized into a test set at least 2000 times. Specification of a positive-valued seed for the pseudorandom number generator ensured that the different factor panels were compared on the same 10 000 dataset partitions. Each sample's collection of test-set prediction probabilities were averaged across iterations to yield a robust, randomization-independent estimate of how the factor panel classifies that sample when it's in a test set. Samples' average test-set prediction probabilities were used to construct an empirical (nonparametric) and binormal (parametric) receiver operating-characteristic (ROC) curve for each factor panel; binormal ROC curves are shown in Fig. 2. For each factor panel, the AUC was calculated for both curve types. ROC-curve estimation was performed using NCSS 2004 (Number Cruncher Statistical Systems, Kaysville, UT, USA); all other statistical analyses were performed using SAS version 9.1.3 (SAS Institute, Cary, NC, USA).

3 Results

3.1 Patient characteristics

Table 1A shows the age and sex distributions among the groups. At the time of sample collection, the median age (range) in years of 47 (16–80) for the Benign group *versus* 49 (25–81) for the Malignant group (WRS $p = 0.34$); the reference Normal samples, 53 (49–67), tended to be older than thyroid patients (WRS $p = 0.053$). The sex distribution was four males and 20 females in the benign group, *versus* seven males and 16 females in the Malignant group (Fisher's exact $p = 0.32$); the thyroid patients had significantly more males than the reference Normals (0 males, 23 females; Fisher's exact $p = 0.012$). Table 1B shows the distribution of pathology reported among the Benign and Malignant cases.

3.2 Serum cytokines and growth factors

As listed in Table 2, concentrations of 32 different serum factors belonging to two functional groups (cytokines/chemokines and growth factors) were evaluated in multiplex assays using xMAP technology. Abbreviations are all defined in Table 2. Ten factors (Eotaxin, HGF, IL-12p40, IL-5, IL-6, IL-8, IP-10, MIP-1 α , TNF-RI, and TNF α) had concentrations values that were within range for all samples. With respect to Benign and Malignant sera ($n = 47$ samples total), four factors (G-CSF, MIG, TNF-RII, and IL-4) had OOR-low values for one or two samples (<5%). IL-15 and IL-13, respectively had seven (15%) and 10 (21%) values that were OOR-low. EGF and IL-1 α , respectively had 14 (30%) and 21 (45%) that were OOR-low. RANTES had seven (15%) values that were OOR-high. These 19 factors met the inclusion criteria for subsequent statistical analysis.

3.3 Univariate analysis of serum analytes

Figure 1 shows, in pg/mL, the distribution of the 19 serum factors forwarded to statistical analysis. Most factors had concentrations ranging approximately between 1 and 1000 pg/mL. Table 3 summarizes the distributions of Fig. 1 using medians and inter-quartile ranges for the Benign, Malignant, and reference Normal groups. Using KW test at 1% α , five factors showed significant differences among groups. For these five factors, the medians in pg/mL for Normal, Benign, and Malignant samples, respectively were 44.58, 3.24, and 2.26 for EGF ($p < 0.001$), 204, 149, and 103 for HGF ($p = 0.002$), 2.19, 2.65, and 2.19 for IL-5 ($p = 0.009$), 3.91, 3.04, and 3.08 for IL-8 ($p = 0.001$), and 20 200, 4800, and 6300 for RANTES ($p < 0.001$). With WRS test at 1% α , none of the 19 factors showed a significant univariate difference between Benign and Malignant samples.

3.4 Multivariate analysis of serum analytes

Table 4 shows the vector coefficients and factor r^2 of the discriminant vector extracted by MANOVA. Factors fell into three groups with relatively clear demarcations, as follows: The first group consisted of IL-8, HGF, MIG, IL-12p40, and IL-5, with r^2 s of 22.2, 19.9, 11.7, 11.6, and 10.1%, respectively. The second group consisted of five more factors (IP-10, IL-6, TNF-RII, MIP-1 α , and RANTES), with individual r^2 s ranging from 4.9 to 3.7%, while the third group consisted of the remaining nine factors, with individual r^2 s ranging from 0.8 to 0.1%. The r^2 s of the first group add up to 75.5% of the Benign–Malignant differentiation.

3.5 Logistic regression with permutation resampling

The r^2 s of Table 4 were used to choose sets of two {IL-8, HGF}, three {IL-8, HGF, and MIG}, and four {IL-8, HGF, MIG, IL-12p40} “multivariately most informative” factors for use as marker panels in logistic regression with permutation resampling. Although Table 4 indicates that IL-5, with r^2 10%, was worthy of inclusion, it was excluded in order to have at least ten samples *per* factor in multivariate analyses on the 47 samples. In each resampling, the logistic

regression modeled the probability that a sample belonged in the Malignant group, and whenever the sample was resampled into a test set, the resulting probability was a prediction probability. Out of 10 000 resamplings *per* panel, the average (SD) number of times a sample was resampled into a test set was 2500.0 (16.17) for Benign samples, and 2173.9 (20.44) for Malignant samples; the minimum number of resamplings into a test set was 2474 for Benign samples and 2144 for Malignant samples. The >2000 prediction probabilities *per* sample were averaged together to yield that sample's average test-set prediction probability, a robust, randomization-independent estimate of how the marker panel classifies that sample when it's in a test set. For each marker panel, the binormal ROC and the AUCs for both curve types were calculated (Fig. 2 and Table 5, respectively).

With IL-5 comprising the best univariate "panel", the empirical ROC AUC was 67.9% and the binormal ROC AUC was 65.1%; the 95% confidence intervals (CIs) for both ROC AUCs encompassed 50%, showing that the performance of IL-5 by itself was not significantly better than random given the modest study sample size of 47. Figure 2 shows that the binormal ROC curve dips below the reference $Y = X$ diagonal for values of 1-specificity below 6%, suggesting that IL-5 by itself might do slightly worse than random under conditions requiring high specificity. With the two-marker set {IL-8, HGF}, the ROC AUCs were 77.2% for the empirical curve and 76.4% for the binormal curve; the 95% CI for both AUCs excluded 50%, showing that the performance of the two-marker panel was significantly better than random at discriminating Malignant samples from Benign samples. With the three-marker panel {IL-8, HGF, MIG}, the ROC AUCs decreased slightly; to 75.7% for the empirical curve and 74.7% for the binormal curve, but the 95% CI for both AUCs still excluded 50%. With the four-marker panel {IL-8, HGF, MIG, IL-12p40}, the ROC AUCs increased to 80.3% for the empirical curve and 81.0% for the binormal curve; for both ROC AUCs, the lower limits of the two-sided 95% CI >60%.

4 Discussion

Thyroid cancer can be treated effectively with high cure rates if detected at an early stage, but up to 20% of thyroid cancer patients are diagnosed with advanced disease. For accurate diagnosis, patients are often subjected to an open, surgical biopsy. Our group has previously demonstrated that multiplex analysis of a cytokine biomarker panel could be utilized for the development of a screening test for head and neck cancer [25]; however, the utility of biomarker panels have not been investigated for detection of thyroid cancer [23,26]. In this study, we used xMAP technology to assay 32 cytokines, chemokines, and growth factors in the pretreatment sera of patients presenting in the clinic with thyroid masses. Nineteen of the 32 assayed factors had measured concentrations within assay range for >50% of the thyroid disease subjects; these 19 factors were evaluated statistically for group differences, and in particular, for their ability to classify thyroid disease subjects as being either benign or malignant.

In univariate analysis, five factors (EGF, HGF, IL-5, IL-8, and RANTES) showed statistically significant differences between thyroid disease patients and the reference Normal group, but no factor showed a statistically significant difference between Malignant and Benign cases. In multivariate analysis; however, the following sets of two {IL-8, HGF}, three {IL-8, HGF, MIG}, and four {IL-8, HGF, MIG, IL-12p40} factors classified test-set samples into Benign and Malignant groups with an accuracy rate significantly better than chance. The number of factors on a multivariate classification panel was not allowed to exceed four, despite the promising high r^2 shown by IL-5. With four factors and 47 samples, there was an average of 11.75 samples *per* factor, which is considered sufficient to avoid over-fitting and insure adequate stability of results [24]. Our embedding the logistic regressions within a permutation algorithm for resampling training and test sets, and using only test-set results for computing ROC curves, added additional robustness to the ROC AUC results. As already noted, AUCs

for both empirical and binormal ROC curves were less for the three-factor set than for the two- and four-factor sets. The third and fourth factor in the addition sequence, MIG and IL-12p40, were highly correlated with each other, moderately correlated with IL-8, and weakly correlated with HGF. The decrease in AUC on addition of MIG to {IL-8, HGF} may thus indicate that MIG and IL-12p40 need to be added to the panel together, not separately.

We were reassured that the results of our study supported previous research suggesting the significance of several biomarkers individually, including HGF, IL-8, IL-12p40, MIG, and IL-5. Other biomarkers may play an important role in the development of thyroid cancer. Previous research demonstrated that RANTES expression, mainly by lymphocytes, may be involved in the maintenance of lymphocytic infiltration and, therefore, be involved in the autoimmune responses of Graves' disease [27]. In our study, serum RANTES levels were significantly lower in thyroid disease patients compared to reference Normals, but not significantly different in Benign *versus* Malignant cases. We had no Graves disease patients in the Benign group yet there may have been some undiagnosed Graves cases in the Normal group, as this is a common thyroid condition.

EGF and VEGF and their receptors (EGFR and VEGFR) have been reported to be over-expressed in follicular thyroid cancer (FTC) [28]. These are potential future targets of small-molecule chemotherapy. In our study, median EGF was more than ten-fold lower in both Benign and Malignant subjects compared to reference Normals, which was unexpected given the role of EGF in promoting thyroid cancer proliferation [29] and in inducing cancer-like gene expression profiles in human thyrocytes [30]. IL-6 appears to play multiple functions in thyroid physiology and disease. Downregulation of IL-6 expression may represent a marker of undifferentiated thyroid carcinoma [31], and in our study no difference was seen in serum IL-6 between any of the three groups. The high level of TNF alpha expression was noted for TNF R2/R7 isoforms in thyroid carcinomas [32]. We observed that patients with thyroid disease tended to have lower TNF alpha than the reference Normals, although the difference was not statistically significant.

Met, the receptor for HGF is over-expressed in approximately 90% of papillary thyroid carcinomas. Met is frequently activated in these carcinomas and may stimulate tumor growth. The abundance of Met expression may differentially regulate cell growth, morphogenesis, and migration in response to HGF [33,34]. Met induces paracrine and autocrine HGF production in papillary thyroid carcinomas. Studies also suggested that HGF stimulation of MET+ tumor cells could be one of the molecular mechanisms involved in the recruitment of dendritic cells in the papillary carcinoma of the thyroid [35]. Our study showed serum HGF to be highest in the reference Normals, intermediate in the Benign cases, and lowest in thyroid malignancies. Since this observation is contrary to other reports, this may be a result of our small sample size or may represent a new insight into the HGF–Met function in thyroid cancer.

The study presented focuses on markers which may be associated with malignant nodular thyroid disease. One example is VEGF which may be relevant in patients that might be treated with anti-angiogenic agents. In addition, detection of cytokines associated with anti-inflammatory or tumor permissive T helper type 2 profiles (such as IL-4, IL-5, or IL-10), may suggest a mechanistic basis for thyroid carcinogenesis [36]. The role of inflammatory thyroid disease in carcinogenesis may result from production of such cytokines and chemokines by thyroid cancer cells of infiltrating immune cells/stromal cells, or both [37]. The involvement of MIG in Hashimoto's thyroiditis and Graves' Disease [38,39], suggest that further studies in patients with inflammatory, as well as neoplastic, thyroid disease are warranted.

The optimal strategy to develop biomarkers for use in early detection, diagnosis, prognosis, and response assessment in cancer has been the subject of much discussion [40–46]. The study

presented shows the feasibility of a diagnostic screening tool that could be used eventually in the clinic in conjunction with FNAB. For thyroid cancer, future studies in this area should concentrate on issues relating to prognosis and response assessment, by examining the longitudinal changes in serum concentrations of these biomarkers and investigating their associations with treatment response, relapse, complications, and survival.

Biomarkers identified in our panel merit further investigation as markers of thyroid cancer. Increasing our understanding of the role of biomarkers in the course of thyroid cancer has a great potential to facilitate the development of new diagnostic and treatment modalities for this disease. We were limited to a four marker panel in our multivariate analysis by our small sample size and the need to keep the number of samples per marker to ten or higher in order to assure validity of results. Larger studies might increase or reduce the significance of these markers. With 100 samples, the maximum panel size could be expanded to as many as ten markers with little risk of overfitting.

In conclusion, we show that analysis of multiple serum biomarkers using xMAP technology is a promising approach for the development of diagnostic assays for patients with thyroid cancer. No single biomarker proved capable of sufficient classification accuracy to permit its widespread use, but two or more markers in combination showed considerable performance improvement over the best single marker. Although statistical analysis indicated that correlation of individual markers with thyroid cancer was modest, a combined biomarker panel showed strong ability to differentiate between Benign and Malignant serum samples. This suggests a potential utility for diagnosis of thyroid cancer, which may aid in preoperative prediction of malignancy in patients with nodular thyroid disease and indeterminate FNA results.

Abbreviations

CI, confidence interval; FNAB, fine needle aspiration biopsy; KW, Kruskal–Wallis; MANOVA, multivariate analysis of variance; OOR, out-of-range; RANTES, regulated upon activation, normally T-expressed and presumably secreted; ROC, receiver operating characteristics; WRS, Wilcoxon rank sum.

Acknowledgements

This work was funded by the James Y. Suen, MD, Endowed Chair (to B. C. S.) for salary support and NIH/ NCI R01 grant CA115902 (to R. L. F.). We thank the anonymous reviewers for their many helpful suggestions.

References

1. Cancer Facts and Figures 2007. American Cancer Society; <http://www.cancer.org>
2. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA* 2006;295:2164–2167. [PubMed: 16684987]
3. Schneider AB, Sarne DH. Long-term risks for thyroid cancer and other neoplasms after exposure to radiation. *Nat. Clin. Pract. Endocrinol. Metab* 2005;1:82–91. [PubMed: 16929376]
4. Nagataki S, Nystrom E. Epidemiology and primary prevention of thyroid cancer. *Thyroid* 2002;12:889–896. [PubMed: 12487771]
5. Mack WJ, Preston-Martin S, Bernstein L, Qian D. Lifestyle and other risk factors for thyroid cancer in Los Angeles County females. *Ann. Epidemiol* 2002;12:395–401. [PubMed: 12160598]
6. Wingren G, Hatschek T, Axelson O. Determinants of papillary cancer of the thyroid. *Am. J. Epidemiol* 1993;138:482–491. [PubMed: 8213752]
7. Basolo F, Fiore L, Pollina L, Fontanini G, et al. Reduced expression of interleukin 6 in undifferentiated thyroid carcinoma: *In vitro* and *in vivo* studies. *Clin. Cancer Res* 1998;4:381–387. [PubMed: 9516926]

8. Gharib H, Goellner JR, Johnson DA. Fine-needle aspiration cytology of the thyroid. A 12-year experience with 11 000 biopsies. *Clin. Lab. Med* 1993;13:699–709. [PubMed: 8222583]
9. Jones MK. Management of nodular thyroid disease. The challenge remains identifying which palpable nodules are malignant. *BMJ* 2001;323:293–294. [PubMed: 11498471]
10. Gharib H, Goellner JR. Fine-needle aspiration biopsy of the thyroid: An appraisal. *Ann. Intern. Med* 1993;118:282–289. [PubMed: 8420446]
11. Greaves TS, Olvera M, Florentine BD, Raza AS, et al. Follicular lesions of thyroid: A 5-year fine-needle aspiration experience. *Cancer* 2000;90:335–341. [PubMed: 11156516]
12. de los Santos ET, Keyhani-Rofagha S, Cunningham JJ, Mazzaferri EL. Cystic thyroid nodules. The dilemma of malignant lesions. *Arch. Int. Med* 1990;150:1422–1427. [PubMed: 2196027]
13. Tee YY, Lowe AJ, Brand CA, Judson RT. Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: A comprehensive literature review. *Ann. Surg* 2007;246:714–720. [PubMed: 17968160]
14. Khalid AN, Quraishi SA, Hollenbeak CS, Stack BC Jr. Fine needle aspiration biopsy vs. ultrasound-guided fine needle aspiration biopsy: Cost-effectiveness as a frontline diagnostic modality for solitary thyroid nodules. *Head and Neck Surgery* 2008;30:1035–1039.
15. Ghoneim C, Soula-Rothhut M, Blanchevove C, Martiny L, et al. Activating transcription factor-1-mediated hepatocyte growth factor-induced down-regulation of thrombospondin-1 expression leads to thyroid cancer cell invasion. *J. Biol. Chem* 2007;282:15490–15497. [PubMed: 17409099]Epub 2007 April 4
16. Borrello MG, Alberti L, Fisher A, Degl'innocenti D, et al. Induction of a proinflammatory program in normal human thyrocytes by the RET/PTC1 oncogene. *Proc. Natl. Acad. Sci* 2005;102:14825–14830. [PubMed: 16203990]Epub 2005 October 3
17. Ruggeri RM, Villari D, Simone A, Scarfi R, et al. Co-expression of interleukin-6 (IL-6) and interleukin-6 receptor (IL-6R) in thyroid nodules is associated with co-expression of CD30 ligand/CD30 receptor. *J. Endocrinol. Invest* 2002;25:959–966. [PubMed: 12553555]
18. Niedźwiecki S, Stepień T, Kopeć K, Kuzdak K, et al. Angiopoietin 1 (Ang-1), angiopoietin 2 (Ang-2) and Tie-2 (a receptor tyrosine kinase) concentrations in peripheral blood of patients with thyroid cancers. *Cytokine* 2006;36:291–295. [PubMed: 17374490]Epub 2007 March 19
19. Vesely D, Astl J, Lastůuvka P, Matucha P, et al. Serum levels of IGF-I, HGF, TGF beta 1, bFGF and VEGF in thyroid gland tumors. *Physiol. Res* 2004;53:83–89. [PubMed: 14984318]
20. Phan HT, Jager PL, van der Wal JE, Sluiter WJ, et al. The follow-up of patients with differentiated thyroid cancer and undetectable thyroglobulin (Tg) and Tg antibodies during ablation. *Eur. J. Endocrinol* 2008;158:77–83. [PubMed: 18166820]
21. Fajarewicz K, Jarzab M, Eszlinger M, Krohn K, et al. A multi-gene approach to differentiate papillary thyroid carcinoma from benign lesions: Gene selection using support vector machines with bootstrapping. *Endocr. Relat. Cancer* 2007;14:809–826. [PubMed: 17914110]
22. Pang S, Smith J, Onley D, Reeve J, et al. A comparability study of the emerging protein array platforms with established ELISA procedures. *J. Immunol. Methods* 2005;302:1–12. [PubMed: 15993890]
23. Sanchez-Carbayo M. Antibody arrays: Technical considerations and clinical applications in cancer. *Clin. Chem* 2006;52:1651–1659. [PubMed: 16809399]Epub 2006 Jun 29
24. van Belle, G., editor. *Statistical Rules of Thumb*. New York: John Wiley & Sons; 2002. p. 221
25. Hathaway B, Landsittel DP, Gooding W, Whiteside TL, et al. Multiplexed analysis of serum cytokines as biomarkers in squamous cell carcinoma of the head and neck patients. *Laryngoscope* 2005;115:522–527. [PubMed: 15744170]
26. Vignali DA. Multiplexed particle-based flow cytometric assays. *J. Immunol. Methods* 2000;243:243–255. [PubMed: 10986418]
27. Simchen C, Lehmann I, Sittig D, Steinert M, Aust G. Expression and regulation of regulated on activation, normal T cells expressed and secreted in thyroid tissue of patients with Graves' disease and thyroid autonomy and in thyroid-derived cell populations. *J. Clin. Endocrinol. Metab* 2000;85:4758–4764. [PubMed: 11134139]
28. Younes MN, Yazici YD, Kim S, Jasser SA, et al. Dual epidermal growth factor receptor and vascular endothelial growth factor receptor inhibition with NVP-AEE788 for the treatment of aggressive follicular thyroid cancer. *Clin. Cancer Res* 2006;3425–3434. [PubMed: 16740767]

29. Hoelting T, Siperstein AE, Clark OH, Duh QY. Epidermal growth factor enhances proliferation, migration, and invasion of follicular and papillary thyroid cancer *in vitro* and *in vivo*. *J. Clin. Endocrinol. Metab* 1994;79:401–408. [PubMed: 8045955]
30. Hébrant A, van Staveren WC, Delys L, Solís DW, et al. Long-term EGF/serum-treated human thyrocytes mimic papillary thyroid carcinomas with regard to gene expression. *Exp. Cell. Res* 2007;313:3276–3284. [PubMed: 17689531]
31. Basolo F, Fiore L, Pollina L, Fontanini G, et al. Reduced expression of interleukin in undifferentiated thyroid carcinoma: *In vitro* and *in vivo* studies. *Clin. Cancer Res* 1998;4:381–387. [PubMed: 9516926]
32. Zubelewicz B, Muc-Wierzoń M, Wierzoń J, Romanowski W, et al. Genetic dysregulation of gene coding tumor necrosis factor alpha receptors (TNF alpha Rs) in follicular thyroid cancer-preliminary report. *J. Biol. Regul. Homeost. Agents* 2002;16:98–104. [PubMed: 12144133]
33. Scarpino S, Stoppacciaro A, Ballerini F, Marchesi M, et al. Papillary carcinoma of the thyroid: Hepatocyte growth factor (HGF) stimulates tumor cells to release chemokines active in recruiting dendritic cells. *Am. J. Pathol* 2000;156:831–837. [PubMed: 10702399]
34. Activation of the hepatocyte growth factor (HGF). Met system in papillary thyroid cancer: Biological effects of HGF in thyroid cancer cells depend on Met expression levels. *Endocrinology* 2004;145:4355–4365. [PubMed: 15192042]
35. Trovato M, Villari D, Bartolone L, Spinella S, et al. Expression of the hepatocyte growth factor and c-met in normal thyroid, non-neoplastic, and neoplastic nodules. *Thyroid* 1998;8:125–131. [PubMed: 9510120]
36. Couch ME, Ferris RL, Brennan JA, Koch WM, et al. Alteration of cellular and humoral immunity by mutant p53 protein and processed mutant peptide in head and neck cancer. *Clin. Cancer Res* 2007;13:7199–7206. [PubMed: 18056201]
37. Linkov F, Lisovich A, Yurkovetsky Z, Marrangoniz A, et al. Early detection of head and neck cancer: Development of a novel screening tool using multiplexed immunobead-based biomarker profiling. *Cancer Epidemiol. Biomarkers Prev* 2007;16:102–107. [PubMed: 17220337]
38. Wang J, Seethala RR, Zhang Q, Gooding W, et al. Autocrine and Paracrine Chemokine Receptor 7 (CCR7) Activation in Head and Neck Cancer: Implications for therapy. *J. Nat. Cancer Inst* 2008;100:502–512. [PubMed: 18364504]
39. Kemp EH, Metcalfe RA, Smith KA, Woodroffe MN, et al. Detection and localization of chemokine gene expression in autoimmune thyroid disease. *Clin. Endocrinol. (Oxf)* 2003;59:207–213. [PubMed: 12864798]
40. Newman, TB.; Browner, WS.; Cummings, SR. Designing Clinical Research. Hulley, SB.; Cummings, SR.; Browner, WS.; Grady, D.; Hearst, N., editors. Philadelphia, PA: Lippincott Williams & Wilkins; 2001. p. 175-194.
41. Maruvada P, Srivastava S. Joint National Cancer Institute-Food and Drug Administration workshop on research strategies, study designs, and statistical approaches to biomarker validation for cancer diagnosis and detection. *Cancer Epidemiol. Biomarkers Prev* 2006;15:1078–1082. [PubMed: 16775163]
42. Baker SG. The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer. *J. Natl. Cancer Inst* 2003;95:511–515. [PubMed: 12671018]
43. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clin. Chem* 2003;49:1–6. [PubMed: 12507953]
44. McShane LM, Altman DG, Sauerbrei W, Taube SE, et al. Reporting recommendations for tumor marker prognostic studies. *J. Clin. Oncol* 2005;23:9067–9072. [PubMed: 16172462]
45. Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J. Natl. Cancer Inst* 2003;95:634–635. [PubMed: 12734304]
46. Somorjai RL, Dolenko B, Baumgartner R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: Curses, caveats, cautions. *Bioinformatics* 2003;19:1484–1491. [PubMed: 12912828]

Figure 1.

Dot plots of serum concentrations (in pg/mL) of the 19 factors subjected to statistical analysis. Green plus signs denote reference Normals; blue circles denote Benign samples; red triangles denote Malignant samples.

Figure 2.

Empirical ROC curves, constructed from *per* patient averages of test-set prediction probabilities generated using logistic regression with permutation resampling. Data were resampled 10 000 times into training sets and test sets for logistic regression using panels of 2–4 factors. For each sample, the >2000 prediction probabilities resulting from when it was resampled into test sets were averaged together, and ROC curves were based on these per-patient averages. The four multivariate most informative markers used in the panels are ranked in order as follows: IL-8, HGF, MIG, and IL-12p40; rankings utilize the marker r^2 of Table 4. IL-5 was the univariately most discriminative marker, and was accordingly used univariately in the analysis.

Table 1

Table 1A. Demographics of study subjects and reference Normals					
Patient emographics		<i>p</i> values ^{a)}			
Patient characteristics	Reference Normals (N = 23)	Benign cases (N = 24)	Malignant cases (N = 23)	Normals versus thyroid disease ^{b)}	Benign versus malignant
Number (%) male	0 (0%)	4 (17%)	7 (30%)	0.012	0.32
Number (%) female	23 (100%)	20 (83%)	16 (70%)		
Patient age (years)					
Median	53	47	49	0.053	0.34
Quartiles	50–57	31–58	42–65		
Full range	49–67	16–80	25–81		

Table 1B. Thyroid conditions among study subjects		
Thyroid pathology		
Pathology report	Benign (N = 24)	Malignant (N = 23)
Goiter and nodular goiter	4 and 13	
Hashimoto's	4	
Thyroiditis	3	
Follicular and hurthle cell		1 and 2
Medullary		1
Insular and papillary ^{a)}		1 and 18

^{a)} *p* values for sex (Fisher's), age (two-sided WRS).

^{b)} Thyroid disease group consists of both benigns and malignants.

^{a)} Papillary, *n* = 11; Papillary/follicular, *n* = 6; Papillary/goiter, *n* = 1.

Table 2

Serum factors names, abbreviations, and functions

Functional group	Serum factor			
	Abbreviation	Full name	Analyzed	Reported function
Cytokine/Chemokine	Eotaxin	Eotaxin-1	X	Attracts eosinophils
	IFN- α	Interferon-alpha	—	Activates macrophages and NK
	IFN- γ	Interferon-gamma	—	Induced by foreign macromolecules
	IL-10	Interleukin-10	—	Reduces cytokine/chemokine production
	IL-12 p40	Interleukin-12 p40 subunit	X	Stimulates NK to produce IFN-g
	IL-13	Interleukin-13	X	Induces IgE
	IL-15	Interleukin-15	X	Induces proliferation of NK
	IL-17	Interleukin-17	—	Proinflammatory
	IL-1 α	Interleukin-1-alpha	X	Stimulates immune, inflammatory, and hematopoietic response
	IL-1 β	Interleukin-1-beta	—	COX2 upregulation
	IL-2	Interleukin-2	—	Stimulates CD4
	IL-4	Interleukin-4	X	B and T cell differentiation
	IL-5	Interleukin-5	X	Promotes B cells and IgA
	IL-6	Interleukin-6	X	Acute phase response, produced by macrophages
	IL-7	Interleukin-7	—	Lymphoid progenitor proliferation
	IL-8	Interleukin-8	X	Recruits PMNs
	IP-10	Interferon gamma inducible protein 10	X	Induced by IFN
	MCP-1	Monocyte chemotactic protein alpha	—	Chemotaxis
	MIG	Monocyte induced gamma interferon	X	Cell migration
	MIP-1 α	Macrophage inflammatory protein 1-alpha	X	Activates granulocytes
	MIP-1 β	Macrophage inflammatory protein 1-beta	—	See above
	RANTES	Regulated upon activation, normally T-expressed and presumably secreted	X	Chemotactic for T cells, eosinophils, basophils, leukocytes
	TNF- α	Tumor necrosis factor alpha	X	Produced by macrophages
	TNF-RI	Tumor necrosis factor Receptor I	X	TNF receptor
	TNF-RII	Tumor necrosis factor Receptor II	X	TNF receptor
Growth or Angiogenic	DR5	Death receptor 5	—	Transduces apoptosis
	EGF	Epithelial growth factor	X	Promotes epithelial cells
	bFGF	Basic fibroblast growth factor	—	Promotes fibroblasts
	G-CSF	Granulocyte colony stimulating factor	X	Promotes granulocytes
	GM-CSF	Granulocyte monocyte colony stimulating factor	—	Promotes granulocytes and monocytes
	HGF	Hepatocyte growth factor	X	Promotes hepatocytes

Functional group	Serum factor			
	Abbreviation	Full name	Analyzed	Reported function
	VEGF	Vascular endothelial growth factor	—	Vasculogenesis and angiogenesis

Table 3

Univariate statistical analysis results for serum factors

Serum factor	Reference normals Median (Quartiles)	Benign subjects Median (Quartiles)	Malignancies Median (Quartiles)	p values ^{d)}	
				KW ^{b)}	WRS ^{c)}
EGF	44.58 (28.04–56.69)	3.24 (0.23–5.16)	2.26 (0.02–5.01)	<0.001	0.706
Eotaxin	20.9 (16.0–33.9)	16.1 (12.9–28.5)	18.3 (11.1–24.1)	0.145	0.848
G-CSF	162 (107–213)	212 (152–251)	236 (125–287)	0.200	0.482
HGF	204 (148.9–248)	149 (99–247)	103 (67–178)	0.002	0.033
IL-12 p40	55.9 (35.0–66.4)	74.8 (46.8–06.6)	59.5 (34.5–84.1)	0.263	0.297
IL-13	3.05 (0.30–6.43)	3.47 (0.24–7.89)	4.27 (2.13–6.76)	0.579	0.410
IL-15	1.04 (0.42–2.14)	0.86 (0.18–1.39)	0.63 (0.14–1.66)	0.536	0.749
IL-1a	0.76 (0.16–7.15)	1.01 (0.16–8.51)	0.76 (0.16–8.67)	0.940	0.772
IL-4	1.47 (0.68–2.37)	1.21 (0.71–2.48)	1.07 (0.62–2.51)	0.869	0.790
IL-5	2.19 (1.88–2.43)	2.65 (2.22–4.42)	2.19 (1.88–2.50)	0.009	0.018
IL-6	5.87 (3.75–8.26)	5.94 (4.46–8.14)	6.48 (5.10–8.16)	0.628	0.413
IL-8	3.91 (3.49–4.19)	3.04 (2.65–3.32)	3.08 (2.71–4.51)	0.001	0.354
IP-10	4.67 (2.97–7.11)	8.91 (5.79–3.34)	8.22 (4.45–1.98)	0.017	0.655
MIG	11.5 (6.1–15.8)	12.6 (9.3–20.2)	13.7 (11.5–23.6)	0.149	0.424
MIP-1a	17.1 (11.4–29.1)	14.4 (10.1–25.2)	18.1 (8.7–27.6)	0.897	0.941
RANTES	20 200 (11 600–0 200)	4800 (3500–700)	6300 (4400–10 300)	<0.001	0.430
TNF-RI	299 (194–355)	338 (251–469)	296 (194–465)	0.339	0.322
TNF-RII	236 (108–383)	368 (233–490)	278 (181–478)	0.082	0.610
TNFα	12.12 (6.56–21.86)	6.87 (5.65–10.26)	7.18 (4.44–11.19)	0.078	0.733

^{a)} p-values in bold are statistically significant at adjusted alpha = 0.01.

^{b)} KW test of overall equality among all three groups.

^{c)} WRS test (two-sided), comparing malignancies to benign subjects only.

Table 4Discriminant Vector from MANOVA on All 19 Serum Factors^{a)}

Serum factors	Vector coefficients ^{b)}	Factor r^2 with discriminant vector
IL-8	0.4708	22.2%
HGF	-0.4460	19.9%
MIG	0.3413	11.7%
IL-12p40	-0.3399	11.6%
IL-5	-0.3170	10.1%
IP-10	-0.2206	4.9%
IL-6	0.2138	4.6%
TNF-RII	0.2008	4.0%
MIP-1a	-0.2008	4.0%
RANTES	0.1929	3.7%
EOTAXIN	-0.0909	0.8%
IL-1a	-0.0883	0.8%
TNFa	-0.0737	0.5%
TNF-RI	0.0700	0.5%
IL-4	-0.0653	0.4%
IL-15	0.0422	0.2%
IL-13	-0.0324	0.1%
G-CSF	0.0316	0.1%
EGF	-0.0129	0.0%

^{a)} Factors were transformed to log₂, and standardized to zero mean and unit variance for extraction of discriminant vector *via* MANOVA.

^{b)} The discriminant vector is normalized to length = 1.00, and points in direction of maximum separation between Benign and Malignant groups.

^{c)} Factor r^2 with discriminant vector are calculated as the squares of the vector coefficients, and add up to 100%. The five factors in the first group (r^2 s>10%; bold text) account for 75.5% of the total r^2 with the gradient of separation between groups.

Table 5Results of logistic regression with permutation resampling, using panels of from two to four markers^{a)}

Marker panel	ROC AUCs ^{b)}	
	Empirical ^{c)} (95% CI)	Binormal ^{d)} (95% CI)
{IL-8 HGF}	0.772 (0.599–0.876)	0.764 (0.596–0.868)
{IL-8, HGF, MIG}	0.757 (0.583–0.865)	0.747 (0.577–0.855)
{IL-8, HGF, MIG, IL-12p40}	0.803 (0.624–0.901)	0.810 (0.651–0.901)
IL-5 (best univariate marker)	0.679 (0.487–0.808)	0.651 (0.472–0.779)

^{a)} Markers eligible for panel membership have the four highest r^2 s in Table 4. For comparison purposes, the best univariate marker (IL-5) was also subjected to logistic regression with permutation resampling.

^{b)} ROC curves were constructed from each sample's average of test-set prediction probabilities derived from the resampling.

^{c)} AUCs of empirical ROC curves are determined from the actual prediction probabilities for each patient.

^{d)} AUCs of binormal ROC curves (shown in Fig. 2) are calculated from fitting binormal-model continuous curves to the patient prediction probabilities.