Molecular analysis of the HPJ-JT syndrome and sporadic parathyroid carcinogenesis

Haven, C.J.

Chapter 8

Gene expression of parathyroid tumors and identification of the potential malignant phenotype.

J Med Genet. 2003 Sep;40(9):657-63.

[CANCER RESEARCH 64, 7405–7411, October 15, 2004]

Gene Expression of Parathyroid Tumors: Molecular Subclassification and Identification of the Potential Malignant Phenotype

Carola J. Haven,[1,3] Viive M. Howell,[1,6] Paul H. C. Eilers,[4] Robert Dunne,[7] Masayuki Takahashi,[1] Marjo van Puijenbroek,[3] Kyle Furge,[2] Job Kievit,[5] Min-Han Tan,[1] Gert Jan Fleuren,[3] Bruce G. Robinson,[6] Leigh W. Delbridge,[8] Jeanette Philips,[9] Anne E. Nelson,[6] Ulf Krause,[10] Henning Dralle,[11] Cuong Hoang-Vu,[11] Oliver Gimm,[11] Hans Morreau,[3] Deborah J. Marsh,[6] and Bin T. Teh[1]

[1]Laboratory of Cancer Genetics and [2]Bioinformatics Special Program, Van Andel Research Institute, Grand Rapids, Michigan; Departments of [3]Pathology, [4]Medical Statistics, and [5]Surgery, Leiden University Medical Center, Leiden, the Netherlands; [6]Kolling Institute of Medical Research, Royal North Shore Hospital and University of Sydney, Sydney, New South Wales, Australia; [7]Commonwealth Scientific and Industrial Research Organisation, Mathematical and Information Sciences, North Ryde, Sydney, Australia; [8]Department of Surgery, Royal North Shore Hospital, New South Wales, Australia; [9]Pacific Laboratory Medicine Services, Royal North Shore Hospital and Department of Pathology, University of Sydney; Departments of [10]Pathology and [11]General, Visceral and Vascular Surgery, Martin Luther University, Halle-Wittenberg, Germany

ABSTRACT

Parathyroid tumors are heterogeneous, and diagnosis is often difficult using histologic and clinical features.

We have undertaken expression profiling of 53 hereditary and sporadic parathyroid tumors to better define the molecular genetics of parathyroid tumors. A class discovery approach identified three distinct groups: (1) predominantly hyperplasia cluster, (2) *HRPT2*/carcinoma cluster consisting of sporadic carcinomas and benign and malignant tumors from Hyperparathyroidism-Jaw Tumor Syndrome patients, and (3) adenoma cluster consisting mainly of primary adenoma and MEN 1 tumors. Gene sets able to distinguish between the groups were identified and may serve as diagnostic biomarkers. We demonstrated, by both gene and protein expression, that *Histone 1 Family 2, amyloid β precursor protein*, and *E-cadherin* are useful markers for parathyroid carcinoma and suggest that the presence of a *HRPT2* mutation, whether germ-line or somatic, strongly influences the expression pattern of these 3 genes. Cluster 2, characterized by *HRPT2* mutations, was the most striking, suggesting that parathyroid tumors with somatic *HRPT2* mutation or tumors developing on a background of germ-line *HRPT2* mutation follow pathways distinct from those involved in mutant MEN 1-related parathyroid tumors. Furthermore, our findings likely preclude an adenoma to carcinoma progression model for parathyroid tumorigenesis outside of the presence of either a germ-line or somatic *HRPT2* mutation. These findings provide insights into the molecular pathways involved in parathyroid tumorigenesis and will contribute to a better understanding, diagnosis, and treatment of parathyroid tumors.

INTRODUCTION

Hyperparathyroidism is a common endocrinopathy, believed to affect ~3 in 1,000 adults (1). It is characterized by calcium-insensitive hypersecretion of parathyroid hormone and increased parathyroid cell proliferation. Hyperparathyroidism may be classified as a primary, secondary, or tertiary disorder. Primary hyperparathyroidism is caused by an inherently abnormal or excessive growth of the parathyroid glands. Secondary hyperparathyroidism (hyperplasia) develops in response to chronic depression of serum calcium levels, generally due to renal impairment. In a minority of patients, this parathyroid hyperactivity becomes autonomous, resulting in tertiary hyperparathyroidism. Hyperparathyroidism may also arise in response to lithium treatment as a therapy for bipolar disorder.

The majority of tumors in primary hyperparathyroidism are sporadic; however, ~5% are associated with the autosomal dominant hereditary cancer syndromes Multiple Endocrine Neoplasia types 1 (MEN 1; OMIM #131100) and 2A (MEN 2A; OMIM #171400), Hyperparathyroidism-Jaw Tumor Syndrome (OMIM #145001), and Familial Isolated Hyperparathyroidism (OMIM #145000).

Histologically, primary hyperparathyroidism can be attributed to a single adenoma in 80% to 85% of cases, hyperplasia in 15% to 20% of cases, and carcinoma in <1% of cases (2). However, parathyroid tumors are heterogeneous, and the differences between histologic types are subtle, confounding the classification. Hyperplasia is defined as an enlargement of more than two glands, whereas adenoma is traditionally considered to be a single gland disorder. However, double or multiple adenomas have been reported in patients with primary hyperparathyroidism including MEN 1 (3), additionally confounding the distinction between adenoma and hyperplasia. Studies of clonality have not been able to clearly distinguish between these tumor types (4), and the diagnosis of parathyroid tumor subtypes remains challenging.

The molecular events involved in the formation of parathyroid lesions are poorly understood. Two genes, *cyclin D1* (*CCND1*) and *MEN1*, have been established as having major roles in parathyroid tumorigenesis. The tumor suppressor gene *MEN1* is involved in the formation of sporadic as well as familial MEN 1 tumors. Recently, we have shown that the putative tumor suppressor gene *HRPT2* is mutated in sporadic parathyroid carcinoma (5) and a small subset of cystic parathyroid adenomas (6), likely playing an important role in the development of these tumors.

Both the *calcium sensing receptor* (*CaSR*) and *vitamin D receptor* (*VDR*) are also known to play a role in parathyroid tumorigenesis (7).

To additionally elucidate the underlying molecular mechanisms involved in the formation of parathyroid lesions and, thus, improve clinical diagnosis and management of these patients, we have undertaken gene expression profiling. We report results for 53 parathyroid tumors belonging to 11 different clinical entities: normal tissue, MEN 1, MEN 2A, Hyperparathyroidism-Jaw Tumor Syndrome, familial isolated hyperparathyroidism, primary hyperparathyroidism, secondary hyperparathyroidism, tertiary hyperparathyroidism, carcinoma, adenoma, and a lithium-associated tumor.

Received 6/16/04; revised 7/27/04; accepted 8/11/04.

Grant support: Dora Lush Postgraduate Research Scholarship, National Health and Medical Research Council Australia; a Cancer Memorial Research Scholarship, Royal North Shore Hospital, Australia; and a Northern Sydney Health Ramsay HealthCare Study Fellowship (V. M. Howell); R. D. Wright Fellowship, National Health and Medical Research Council, Australia (D. J. Marsh); in part the Pinguin-Stiftlung (C. Hoang-Vu); and the Van Andel Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: C. J. Haven and V. M. Howell contributed equally to this work. A. Nelson is currently at the Garvan Institute of Medical Research, Darlinghurst, Sydney, Australia. Supplementary data regarding the Gene-Rave analysis, including details of normalization, treatment of missing spots, pairwise class discrimination error matrices, and histograms of cross-validation error rates are available to download from http://www.kolling.usyd.edu.au or http://www.bioinformatics.csiro.au. Supplementary Tables S1-S6 can be found at Cancer Research Online (http://cancerres.aacrjournals.org).

Requests for reprints: Debbie Marsh, Kolling Institute of Medical Research, Royal North Shore Hospital, St. Leonards, New South Wales, Australia, 2065. E-mail: Debbie_Marsh@med.usyd.edu.au; Hans Morreau, Department of Pathology, Leiden University Medical Center, Albinusdreef 2, PO Box 9600, Building 1, LI-Q 2300 RC Leiden, The Netherlands. E-mail: J.Morreau@lume.nl; Bin Teh, Van Andel Research Institute, 333 Bostwick Ave NE, Grand Rapids, MI 49503. E-mail: Bin.Teh@vai.org.

MATERIALS AND METHODS

Gene Expression Profiling

Tumor Samples. A total of 53 parathyroid tumors and 16 normal specimens of parathyroid tissue were obtained from the Leiden University Medical Center, Royal North Shore Hospital, and Martin Luther University. Normal parathyroid tissue was pooled from excess cell washings after routine parathyroid autotransplantation during thyroidectomy. Approval for this study was obtained from the Human Research Ethics Committees of the participating institutions and the Van Andel Research Institute. Tumors were snap-frozen in liquid nitrogen immediately after surgery and stored at −80°C until use. Patient data are summarized in Supplementary Data Table S1. Histologic classification was established according to WHO published guidelines (8).

Fifteen familial tumor specimens were collected. Five were from two Hyperparathyroidism-Jaw Tumor Syndrome families. Two of these 5 tumors were classified as carcinoma and 3 as adenoma. All harbored germ-line *HRPT2* mutations. Five tumors (2 single adenomas and 3 multigland hyperplasia: Table S1) were from MEN 1 families and had documented *MEN1* mutations. One tumor was from a clinically diagnosed MEN 2A patient, and 4 were from familial isolated hyperparathyroidism families. No additional tumors were found in the familial isolated hyperparathyroidism families. Two familial isolated hyperparathyroidism patients harbored a germ-line *HRPT2* mutation. Linkage to the *MEN1* locus 11q13 could not be excluded in the other two familial isolated hyperparathyroidism families, and no germ-line *MEN1* mutations have been detected. Of the sporadic tumors, 16 were classified as adenomas, 1 as lithium-associated, 5 as carcinoma, and 16 as hyperplasia. The hyperplasia specimens were additionally classified as primary ($n = 2$), secondary ($n = 9$), or tertiary ($n = 5$). The percentage of neoplastic tissue in each sample was assessed histologically to ensure the presence of at least 70% neoplastic cells. However, the possibility of contamination of parathyroid tissue with adjacent nonparathyroid tissue could not entirely be excluded due to wide resection of the carcinoma tumors.

Preparation of RNA. Total RNA was isolated from frozen tissues using TRIzol reagent (Invitrogen, Carlsbad, CA) and purified either by precipitation with 2.5 mol/L $LiCl_2$ or Qiagen RNeasy spin columns (Qiagen Inc., Valencia, CA) according to the manufacturer's protocols.

cDNA Microarray Fabrication and Experimental Procedure. Microarray slides spotted with 19,968 cDNA clones from the Research Genetics 40K Human Clone Set (Research Genetics Inc., Huntsville, AL) were fabricated at the Van Andel Research Institute as described previously (9), and microarray experiments were performed as described previously (10). Briefly, 25 to 50 μg of total RNA from parathyroid tissue and an equal quantity of Universal Human Reference total RNA (Stratagene, Cedar Creek, TX), were reverse transcribed using Superscript II (Invitrogen) and oligo $d(T)_{20}VN$ in the presence of Cy5-dCTP and Cy3-dCTP (Amersham Inc., Piscataway, NJ), respectively. After direct labeling, the two probes were hybridized for 20 hours at 50°C to a microarray slide. The slides were then washed, immediately dried, and scanned at 532 nm and 635 nm in a Scan Array Lite (Perkin-Elmer Life, Boston, MA). Reciprocal labeling was performed on 3 samples (Supplementary Table S1). All of the other samples were tested in singlicate.

Data Analysis. Images were analyzed using Genepix Pro 3.0 (Axon, Union City, CA). The ratio of Cy5 intensity to Cy3 intensity for each spot represented tumor RNA expression relative to the Universal Human Reference. The Genepix result files were individually normalized using Another Microarray Database based on spots with pixel correlation values >0.75. Flagged spots and spots with background-subtracted intensities <150 in either the Cy5 or Cy3 channel were excluded from cluster analysis.

To check for color bias and dye incorporation efficiency, the correlation coefficient (r) was determined for each reciprocal-labeled array against its matched standard-labeled array. In all three cases $r > 0.8$, the results were averaged, and these averages were used for additional analysis. In the pooled normal sample all of the genes for which two results were not available or where the difference between the results was outside the 95% confidence interval of the log-transformed ratio, results were excluded from additional analysis. For some analyses, the Cy5:Cy3 ratios for the remaining genes in the tumor arrays were divided by the averaged normal result so that the ratios now represented tumor expression relative to normal parathyroid tissue.

Gene expression values were visualized using CLUSTER and TREEVIEW software (11). Ratios were log-transformed and median-centered so that the median log-transformed ratio equalled zero. The gene expression ratios were median-polished across all of the samples. To test the influence of different gene sets on the unsupervised clustering, the filtering parameters (genes present: 80% to 100%; number of observations: 2 to 20; fold change: 2 to 4) in CLUSTER were varied to create distinct gene sets.

Statistical Classification and Supervised Cluster Analysis Using Cluster Identification Tool and Significance Analysis of Microarrays. Cluster Identification Tool software was used to find genes that were differentially expressed (using Student's *t* test) between the clusters (12). To find significantly discriminating genes, 1,000 t-statistics were calculated by randomly placing patients into two groups (for example, cluster 1 *versus* the remainder). A 99.9% significance threshold ($\alpha \leq 0.01$) was used to identify genes that could significantly distinguish between two clusters *versus* the random patient groupings. In addition, Significance Analysis of Microarrays was used to identify genes with statistically significant changes in expression by assimilating a set of gene-specific *t* tests (13).

Statistical Classification and Probability Scores in Clinical Diagnostic Categories by Penalized Logistic Regression and Significance Analysis of Microarrays. We used a Penalized Logistic Regression model (14) in a two-step approach to identify molecular classifiers for each cluster. The first step separated clusters 1 and 3 from cluster 2. In the second step (including only tumors that had a low probability of belonging to cluster 2), cluster 3 samples were separated from the remainder. Penalized Logistic Regression generated a probability score for each array, indicating the chance for a sample to be part of one of two *a priori* defined groups. These probabilities were computed from the logarithmic expression of a selected set of genes on an array. This selection was not based on a predictive measure of the performance of a gene but simply on it being measured reliably in all arrays. To decrease the set of genes, we used Significance Analysis of Microarrays to rank the genes. We performed a Penalized Logistic Regression analysis with different numbers of genes to find the gene set with the smallest number that could still predict robustly. An acceptable prediction was performed using a maximum of 1,460 genes (genes with results for all arrays) or a minimum of 50 genes for each step.

The model involved a penalty parameter. It was not automatically optimized with Akaike's information criterion but rather was set manually to a value that gave log-odds (base 10) in the approximate range of −3 to 3. To determine the variability of the estimated log-odds, 100 bootstrap samples were taken and means and SDs computed.

Statistical Classification Using Gene-Rave. The expression array data were independently analyzed to find small gene sets suitable as accurate molecular classifiers for the different clinical groups. For this analysis the array results were normalized using a series of strategies including loess fitting and spatial smoothing (15, 16). After normalization, flagged spots with $M > 5$, where $M = \log_2 (Cy5/Cy3)$, were removed. The data were then analyzed by pair-wise discrimination of clinical groups using Gene-Rave software. Only groups with array results for at least 5 tumors were analyzed, *i.e.,* Hyperparathyroidism-Jaw Tumor Syndrome, sporadic carcinoma, sporadic adenoma, hyperplasia, and MEN 1. Gene-Rave uses a Bayesian penalty term with an improper prior that puts a high weight on finding a small solution set of genes. Cross-validation and permutation testing are used to assess the prediction error rate (17). A canonical variate plot produced by the "profile analysis" algorithm was used to display the data in low dimensional space. This algorithm involves a factor analytical approach for the within-class covariance matrices specifically for use with large numbers of variables (genes).

Statistical Analyses. χ^2 analysis (SPSS version 10.0) was used to determine the significance of associations between RNA (as measured by expression microarray) and protein expression and correlations between expression profiles and cluster groups.

Immunohistochemical Analysis of Protein Expression

Samples. Paraffin-embedded blocks from 149 parathyroid tumors (16 of which also had cDNA microarrays performed) and 9 normal parathyroid tissues were used for the construction of a tissue microarray. Eighty-seven patients were diagnosed as having primary parathyroid adenomas (80 sporadic, 5 MEN 1-related, and 2 Hyperparathyroidism-Jaw Tumor Syndrome -related), and 26 were diagnosed with hyperplasia (12 with primary hyperparathyroid-

ism, 3 with secondary hyperparathyroidism, 3 with tertiary hyperparathyroidism, and 8 MEN 2A-related). In addition to these benign tumors, 36 tumors from 30 patients with parathyroid carcinoma were analyzed.

Immunohistochemical Procedures. A paraffin sectioning aid system (Instrumedics Inc., Hackensack, NJ) was used to facilitate cutting of 5-μm sections of the tissue microarray. The sections were deparaffinized, treated with 0.3% H_2O_2 in CH_3OH, and submitted to antigen retrieval (microwave oven treatment for 10 minutes in 10 mmol/L citrate buffer pH 6), except histone H1, for which no antigen retrieval was necessary. Tissue sections were incubated overnight at room temperature with mouse antihuman amyloid βA4 precursor protein (APP; dilution 1:80, clone 22C11, Boehringer, Ingelheim GmbH), H1 (dilution 1:6400, clone 1415-I, Neomarkers, Fremont, CA), cyclin D1 (CCND1; dilution 1:500, clone DCS-6, Neomarkers, Fremont, CA), and E-Cadherin (CDH1; dilution 1:500, clone HECD1, Zymed, Carlton, CT). Sections were subsequently washed (3 × 5 minutes in PBS) and incubated (30 minutes) with biotinylated secondary antibody in PBS/bovine serum albumen 1%, washed (3 × 5 minutes in PBS) and incubated (30 minutes) with horseradish peroxidase-streptavidin complex. Diaminobenzidine tetrahydrochloride was used as a chromogen followed by counterstaining with hematoxylin. The primary antibody was omitted as a negative control. Brain (sclerotic plaques) and cervix served as positive controls for APP and CDH1, respectively, and tonsil was used for CCND1 and H1. Expression was scored by light microscopy. The cutoff parameters used for classification of normal and overexpression are summarized in Table 1.

RESULTS AND DISCUSSION

Gene Expression Profiling

Unsupervised Clustering. Using expression microarray analysis of 11 clinical parathyroid entities, we have identified three broad and distinct tumor groupings based on unsupervised clustering according to gene expression profiles. Using different gene sets by varying the filtering parameters, these three clusters remained constant, although the relative positions of the clusters did vary. The dendrogram in Fig. 1 is derived from hierarchical clustering with a set of 6,150 genes. The composition of the three clusters reflected, in part, the histologic classification of the tumors. However, concordance with histologic classification was not complete. Cluster 1 was composed of predominantly hyperplastic specimens but also a lithium-associated tumor, a MEN 2A adenoma, and 3 sporadic adenomas. The pooled normal sample also clustered with this group.

The most robust cluster identified in this study was cluster 2, which was composed of the sporadic carcinomas, familial Hyperparathyroidism-Jaw Tumor Syndrome (both benign and malignant tumors), and 2 familial isolated hyperparathyroidism tumors. Two sporadic carcinomas in the small cluster 2 subgroup (#779g and #1798g) demonstrated expression of some thyroid-specific genes (including *metallothionein 1H, E,* and *G*), suggesting possible admix with nearby thyroid tissue that may have occurred due to extensive surgical clearance. This was not observed in other tumors in this cluster, and for this reason these 2 tumors were excluded from additional analyses. The small intracluster variance (V = 0.25) and the large distance from the other two clusters demonstrated the distinct nature of the gene expression profile

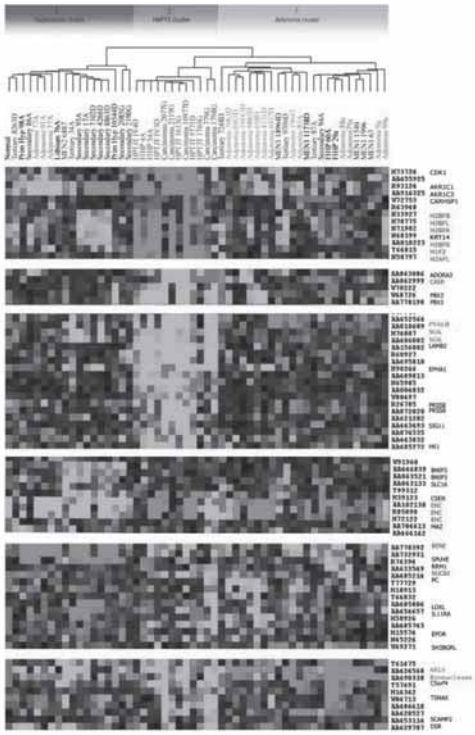

Fig. 1. Detailed sample dendrogram of unsupervised hierarchical clustering. *Red* and *green* indicate transcript expression levels respectively above and below the median (*black*) for each gene across all samples. *Grey squares* indicate no results.

of this cluster. These results, in combination with the knowledge that all of the tumors tested in this cluster (11 of 12) harbored *HRPT2* mutations (Supplementary Data Table S1), support a distinct molecular pathway of parathyroid tumorigenesis for sporadic carcinomas, familial Hyperparathyroidism-Jaw Tumor Syndrome and familial isolated hyperparathyroidism tumors with a *HRPT2* mutation, that is separate from that for other parathyroid tumors. Our results are striking in that we demonstrate that Hyperparathyroidism-Jaw Tumor Syndrome adenomas, otherwise indistinguishable from non-Hyperparathyroidism-Jaw Tumor Syndrome adenomas, exhibit a common microarray signature with parathyroid carcinomas, which have a high

Table 1 *Cut-off values*

Molecule	Method	Normal	Overexpression
E-cadherin	IHC	Membranous staining/no staining	Irregular membranous staining or depositions/droplets in cell
	MA	Fold change <2	Fold change >2
Amyloid βA4 precursor protein	IHC	Weak/moderate cytoplasmic staining	Strong cytoplasmic staining
	MA	Fold change <2	Fold change >2
Histone H1	IHC	≤60% of positive nuclei	≥60% of positive nuclei
H1F2	MA	Fold change <2	Fold change >2
Cyclin D1	IHC	<10% positive cells	>30% positive cells
	MA	Fold change <4	Fold change >4

NOTE. MA values are ratios of tumor to normal.
Abbreviations: IHC, immunohistochemistry; MA, cDNA microarray.

incidence of *HRPT2* mutations. These results suggest that *HRPT2* mediates novel and fundamental pathways of tumorigenesis and malignant transformation. This has also been supported by the multiple genetic changes identified by previous comparative genomic hybridization studies of parathyroid carcinoma (18). The increased incidence of parathyroid carcinoma in Hyperparathyroidism-Jaw Tumor Syndrome patients and our demonstration of a close relationship between sporadic carcinomas and familial Hyperparathyroidism-Jaw Tumor Syndrome tumors suggest that apparently benign tumors within this cluster may have the potential to progress to malignant tumors.

Cluster 3 contained the majority of the sporadic adenoma specimens and all of the MEN 1 and 2 familial isolated hyperparathyroidism tumors. Three of 5 tumors, clinically classified as tertiary hyperplasia, as well as 1 secondary hyperplasia, were found in this cluster. This clustering demonstrates at the molecular level the recognized clinical transition from secondary to tertiary or autonomous disease. Additional studies are warranted to confirm this observation and to identify gene sets that separate secondary from tertiary hyperplasia.

All 5 of the MEN 1 tumors (2 single adenomas and 3 multigland hyperplasia) were also located in cluster 3, suggesting that familial MEN 1 tumors and a subset of sporadic adenomas may share a similar genetic pathway of tumorigenesis. This is supported by previous findings of frequent LOH at 11q13 and of *MEN1* mutations in 20% to 40% of sporadic adenomas in contrast to hyperplasia (19).

Two of the 4 familial isolated hyperparathyroidism tumors in this study also clustered with the sporadic adenomas and MEN 1 tumors in cluster 3. These tumors were from families where linkage to the *MEN1* region could not be excluded. The separation of these 2 familial isolated hyperparathyroidism tumors (cluster 3) from the 2 familial isolated hyperparathyroidism tumors harboring *HRPT2* mutations (cluster 2) again clearly indicates that different molecular pathways are involved in familial isolated hyperparathyroidism patients and adds additional weight to a subset of familial isolated hyperparathyroidism being variants of either Hyperparathyroidism-Jaw Tumor Syndrome or MEN 1 rather than a distinct entity. It is possible that those familial isolated hyperparathyroidism tumors displaying a *HRPT2*-like expression profile may benefit from more aggressive clinical management than those displaying a *MEN1*-like profile.

Genes with established roles in parathyroid tumorigenesis were evaluated by comparing the expression levels of these genes in tumors with the expression level in normal parathyroid tissue. The gene encoding parathyroid hormone was highly expressed resulting in image-saturation, thus data unable to be included in additional analysis. *VDR* was down-regulated in all of the groups. *CCND1, parvalbumin (PVALB),* and *CaSR* were differentially expressed with *CCND1* up-regulated in cluster 1 and 2 and *PVALB* down-regulated in cluster 2. *CaSR* was down-regulated in all of the clusters but significantly down-regulated in cluster 2 ($P < 0.05$). There was a trend toward down-regulation of *MEN1* in *MEN1*-related tumors (from patients with MEN 1 or with Familial isolated hyperparathyroidism where linkage to *MEN1* could not be excluded and from sporadic tumors with somatic *MEN1* mutations); however, this did not reach significance, possibly due to the small size of this group ($n = 9$). *TP53* was not differentially expressed in any of the clusters. *RET* and *HRPT2* were not represented on the arrays.

Molecular Classifiers

Cluster Identification Tool. Supervised clustering using Cluster Identification Tool identified 22, 329, and 17 genes differentially expressed between clusters 1, 2, and 3 respectively (Tables S2, S3 and S4, respectively, in the Supplementary Data) In cluster 1, 22 genes

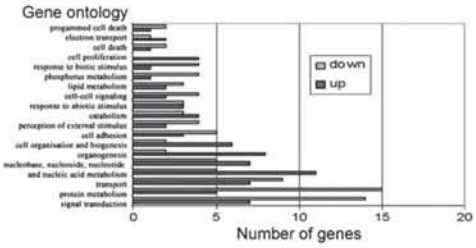

Fig. 2. Gene ontology for biological processes of 204 genes differentially expressed in cluster 2 (Hyperparathyroidism-Jaw Tumor Syndrome/sporadic carcinoma). Only categories with at least 3 genes over- or underexpressed are listed. The ontology information was compiled from GeneCards, Genatlas and DAVID.

were differentially expressed. Of note, *ectodermal-neural cortex* (*ENC1*), *hypothetical protein MGC11034,* and *Homo sapiens clone MGC:16152* were down-regulated relative to the other two groups. Up-regulated were *BENE protein* (*BENE*), *membrane glycoprotein MRC OX-2* (*MOX2*), and *immunoglobulin superfamily member 4* (*IGSF4*).

In cluster 2, of the 329 genes identified by Cluster Identification Tool ($P < 0.05$), 204 were also identified as differentially expressed by Significance Analysis of Microarrays. This subset of genes was classified according to gene ontology information (Fig. 2), which indicated that major changes had occurred to genes involved in signal transduction, protein metabolism, transport, nucleic acid metabolism, organogenesis, cell organization, and cell adhesion. Examination of individual genes in this subset suggests important roles for *KIAA1376* (signal transduction), *Serum/glucocorticoid regulated kinase* (*SGK*; Phosphate metabolism, transport), *Cullin 5* (*CUL5*; cell death), and *Nucleobindin* (*NUCB2),* all of which are down-regulated. Up-regulated are *amyloid β precursor protein* (*APP*; programmed cell death), *E-cadherin* (*CDH1*; cell adhesion and signal transduction), *histone H1 family member2* (*H1F2*; cell organization and biogenesis), the *aldo-keto reductase family 1 C3* (*AKR1C3*; cell proliferation and lipid metabolism), *CD24* (signal transduction), *ubiquitin carboxyl-terminal esterase L1* (*UCHL1*; catabolism), *development and differentiation enhancing factor 1* (*DDEF1*), and *laminin β1* (*LAMB1*; cell adhesion).

In cluster 3, 17 genes were differentially expressed. *Aldehyde dehydrogenase 1 family, member A2* (*ALDH1A2*), and *metallocarboxypeptidase* (*CPX-1*) were down-regulated, and the *KIAA0435 gene product*, *ADP-ribosylation factor-like 5* (*ARL5*), and *exonuclease NEF-sp* were up-regulated relative to the other 2 clusters.

Statistical Classification and Probability Scores in Clinical Diagnostic Categories by Penalized Logistic Regression and Significance Analysis of Microarrays. Using Penalized Logistic Regression analysis, a two-step statistical model for classifying tumors was built (Fig. 3). With 50 genes (Supplementary Data Table S5), Penalized Logistic Regression separated cluster 2 tumors from the rest with a minimal log-odds separation of 1:1,000 between the 2 groups of tumors (Fig. 3*A*), confirming the results obtained by unsupervised hierarchical clustering (Fig. 1). A second step with a different set of 50 genes (Supplementary Data Table S6) separated cluster 3 tumors (predominantly adenomas) from the rest, which now consisted of all cluster 1 specimens (predominantly hyperplasia; Fig. 3*B*). This second-step classification was less pronounced (log-odds separation of 1:100), and larger SDs were apparent, again reflecting the results of hierarchical clustering. Cross-validation confirmed the above separation. The subgroupings observed under clusters 1 and 3 were not apparent in the Penalized Logistic Regression model.

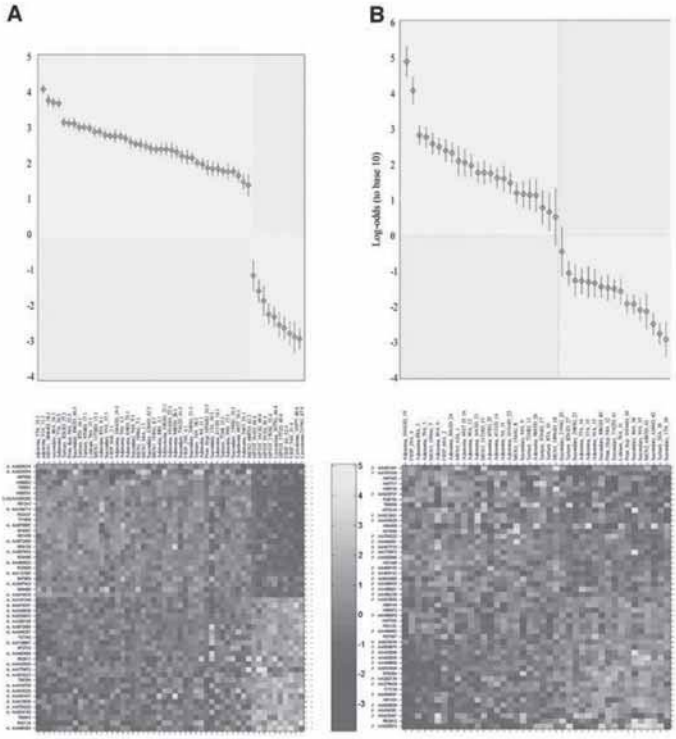

Fig. 3. Results of classification with a penalized logistic regression model, based on two sets of 50 most significant genes (according to Significance Analysis of Microarray) present in all arrays. *Top panel* depicts the results of 100 bootstrap replications: means ([circle]) ±1 SD (*lines*) of log-odds per array. The vertical scale is log-odds to base 10, where "0" represents equal odds, so that in *A*, "2" indicates a probability of 100 to 1 that the specimen is not a cluster 2 tumor, and "−3" alternately indicates a probability of 1000 to 1 that the specimen is a cluster 2 tumor. Outcomes in the green (*pink*) regions are correctly (incorrectly) classified. *A* represents step-1 differentiation between non-Hyperparathyroidism-Jaw Tumor Syndrome/carcinoma tumors (clusters 1 and 3) and Hyperparathyroidism-Jaw Tumor Syndrome/carcinoma tumors (cluster 2) using 50 genes. *B* represents step-2 differentiation between adenomas (cluster 3) and the remainder (cluster 1) using a different set of 50 genes. *Bottom panel* provides graphical presentation of the log expressions with the arrays in the same order as in the *top panel*. The genes have been ordered along the values of their coefficients in the model with *yellow squares* indicating the highest expression.

Statistical Classification Using Gene-Rave. Using a different approach to normalization and analysis of the array data, Gene-Rave confirmed separation of the tumors into the three main clusters as demonstrated by unsupervised clustering (Figs. 1 and 4). When testing for genes to additionally separate the Hyperparathyroidism-Jaw Tumor Syndrome (including the 2 familial isolated hyperparathyroidism specimens, #4A and #54A, with *HRPT2* mutations) and carcinoma groups, the cross-validation error rates for correct classifications were indistinguishable from the permuted versions (Supplementary Data Fig. 15). This indicated that these two classes were inseparable. The Hyperparathyroidism-Jaw Tumor Syndrome/carcinoma, adenoma, hyperplasia, and MEN 1 groups were found to be separable, although the separation between adenoma and MEN 1 was small (Fig. 4). Two sporadic carcinomas (#779G and #1798G) were consistent outliers, as also noted in the unsupervised cluster analysis (Fig. 1) and were excluded from additional analysis. An advantage of Gene-Rave technology is that the gene sets able to discriminate between groups are typically very small. In this study, 9 genes [*UCHL1, CD24, ALDH1A2, PVALB, clone DKFZp434E03, Vascular Cell Adhesion Molecule 1*(*VCAM1*), *Testican-3, MOX2,* and *glutamate decarboxylase 1* (*GAD1*)] were able to discriminate between the four groups.

Immunohistochemical Analysis of Protein Expression

We selected four highly differentially expressed genes for which antibodies to their respective proteins were available, to compare transcript and protein overexpression. Three of these genes, *CDH1, H1F2*, and *APP* were highly overexpressed in cluster 2 compared with clusters 1 and 3. *CCND1* demonstrated differential expression between clusters 1 and 3, with higher expression in cluster 3.

Parameters were established for classification as normal or overexpression of each transcript and protein (Table 1), and examples of the differential expression assessed by immunohistochemical staining are shown in Fig. 5.

Comparing microarray expression with immunohistochemistry in 16 specimens represented on both the expression microarrays and tissue microarray, 94% correlation was found for H1, APP, and CCND1, and 75% for CDH1. To additionally evaluate the expression of these four proteins, all of the tissue microarray tumor specimens were grouped in accordance with the cluster sets created by unsupervised clustering (Fig. 1). For CDH1, H1, and APP, protein overexpression occurred predominantly in the carcinoma and Hyperparathyroidism-Jaw Tumor Syndrome familial tumors, resulting in a significant correlation between the immunohistochemistry and expression data ($P < 0.01$; Fig. 6).

Increased expression of CCND1 is known to play a primary role in the formation of parathyroid tumors. Overexpression of CCND1 was noted primarily in the hyperplasia and carcinoma/Hyperparathyroidism-Jaw Tumor Syndrome tumors, with only 12% of adenomas/MEN 1 tumors demonstrating overexpression by immunohistochemistry. This allowed moderate discrimination between adenoma and hyperplasia and also correlated with the expression microarray results.

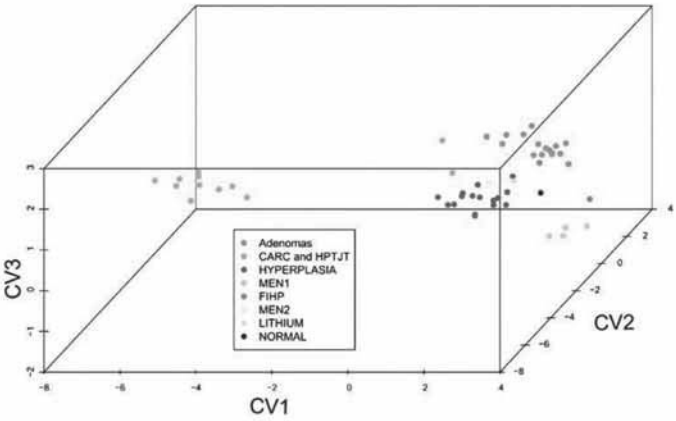

Fig. 4. Canonical variate plot of all arrays. The plot shows the first three canonical variates (*CV1–3*). CV1 consisted of 125 genes, CV2, 57 genes, and CV3, 63 genes. Among genes of significance in CV1–3 were *CDH1, APP, UCHL1, IGSF4, MOX2,* and *GAD1*. A large separation between the carcinoma/Hyperparathyroidism-Jaw Tumor Syndrome group and the rest of the tumors is evident. The 2 familial isolated hyperparathyroidism specimens with *HRPT2* mutations are included in the carcinoma and Hyperparathyroidism-Jaw Tumor Syndrome group that is depicted in green. The 2 (*green*) carcinomas distant from the main Hyperparathyroidism-Jaw Tumor Syndrome/carcinoma cluster are the outliers #779G and #1798G. Clear separation of the adenoma (*red*), hyperplasia (*purple*), and MEN 1 (*light blue*) groups is also evident. Two MEN 1 tumors are overlaid in this analysis and appear as *one blue spot*. The pooled normal (*black*) is among between the adenoma, hyperplasia, and MEN 1 groups. The 2 familial isolated hyperparathyroidism tumors with linkage at 11q13 (*pink*) are located between the adenomas and MEN 1 tumors. The MEN 2A (*yellow*) and lithium-associated tumor (*gray*) are situated closest to the hyperplasia group.

These results are in agreement with previous immunohistochemistry studies in which overexpression of CCND1 has been observed in 50% to 91% of parathyroid carcinomas, up to 61% of sporadic parathyroid hyperplasia, and 18% to 40% of sporadic parathyroid adenomas (20, 21).

Up-regulation of two of four members of the histone family (*H1* and *H2*; Supplementary Data Table S5) was evident in cluster 2 from the microarray experiments, and H1 protein overexpression was confirmed in sporadic carcinomas by an increased percentage of nuclei expressing this histone. It has been postulated recently that H1 may regulate specific genome repair mechanisms in some species (22). Because DNA repair pathways are involved in DNA maintenance, they are likely to have relevance to tumorigenic processes. APP is a functional neuronal receptor, shown in this study to be up-regulated at both the RNA and protein level. Both transcript and protein were found overexpressed in cluster 2. The soluble N-terminal ectodomain (sAPP) is a product of the APP protein that can function as an epithelial growth factor in thyroid cells (23) and has been proposed recently to have a pro-proliferative role in some pancreatic cancer cell types.

CDH1 is a calcium-dependent cell-cell adhesion glycoprotein. Loss of function due to mutation or methylation is thought to contribute to cancer progression by increased proliferation, invasion, or metastasis. It has also been reported that increased parathyroid hormone or low extracellular calcium affects the expression and localization of this molecule (24). We found *CDH1* up-regulated at the mRNA level in cluster 2, with aberrant staining noted by, indicating loss of function in cell adhesion (25).

We also evaluated the diagnostic sensitivity of combined immunohistochemistry using these four proteins. If two or more of these four proteins are overexpressed in a single case, there is an 81% chance that it is a carcinoma, and a 2% and 11% chance that it is an adenoma

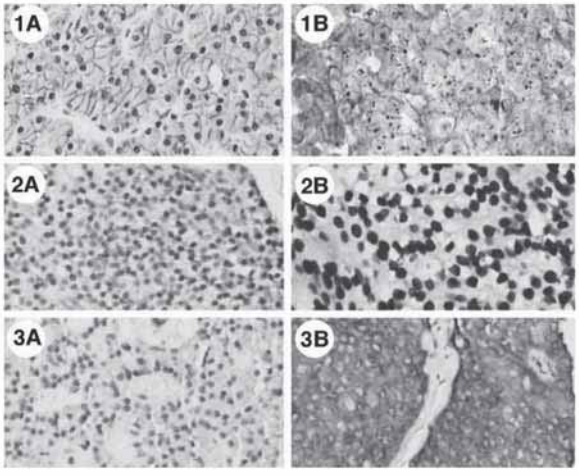

Fig. 5. Differential immunohistochemical staining at ×200 magnification between (*A*) adenomas and (*B*) carcinomas (overexpression) for E-cadherin (*1*), histone H1 (*2*), and amyloid βA4 precursor protein (*3*).

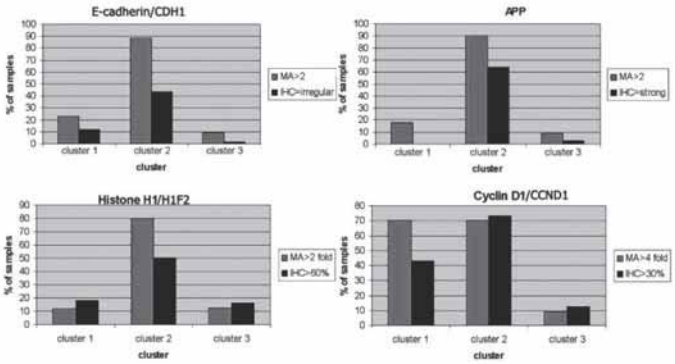

Fig. 6. Correlation between Microarray (*MA*; $n = 53$) and Immunohistochemistry (*IHC*; $n = 147$) data for the three different parathyroid subgroups as defined by unsupervised clustering of the microarray data. The percentage of samples displaying overexpression (as defined in Table 1) in each cluster group are graphed.

or hyperplasia, respectively ($P < 0.01$). We have demonstrated that *H1F2, APP,* and *CDH1*, identified from our microarray experiments, are also useful immunohistochemistry makers for the diagnosis of parathyroid carcinoma. Our results suggest that the presence of a *HRPT2* mutation, whether germ-line or somatic, strongly influences the expression pattern of these 3 genes, and we postulate that these 3 genes may have potential biological significance in parathyroid tumorigenesis.

In conclusion, we have identified three broad cluster groupings of parathyroid disease based on distinct molecular signatures. Cluster 2 was the most striking and included both benign and malignant tumors from patients with Hyperparathyroidism-Jaw Tumor Syndrome, as well as sporadic parathyroid carcinomas with somatic *HRPT2* mutation and a subset of tumors from patients with familial isolated hyperparathyroidism. This cluster is strongly suggestive that parathyroid tumors with somatic *HRPT2* mutation or tumors developing on a background of germ-line *HRPT2* mutation follow pathways distinct from those involved in mutant MEN 1-related parathyroid tumors. Furthermore, our findings likely preclude an adenoma to carcinoma progression model for parathyroid tumorigenesis outside of the presence of either a germ-line or somatic *HRPT2* mutation.

ACKNOWLEDGMENTS

We thank the VARI cDNA microarray facility for the printed microarrays; Tom van Wezel and David Nadziejkan for critically reviewing this manuscript; S. Scollon and H. Dang from VARI for assistance with the microarray analysis; A-L. Richardson, L. Cheung, and G. Theodosopoulos, from the Kolling Institute of Medical Research, for their assistance with DNA extractions; and endocrinologist J. W. A. Smit from Leiden University Medical Center, Department of Endocrinology, for assistance with the collection and clinical classification of tumors.

REFERENCES

1. Adami S, Marcocci C, Gatti D. Epidemiology of primary hyperparathyroidism in Europe. J Bone Miner Res 2002;17 Suppl 2:N18–23.
2. Marx SJ. Hyperparathyroid and hypoparathyroid disorders. N Engl J Med 2000;343: 1863–75.
3. Harness JK, Ramsburg SR, Nishiyama RH, Thompson NW. Multiple adenomas of the parathyroids: do they exist? Arch Surg 1979;114:468–74.
4. Tominaga Y, Kohara S, Namii Y, et al. Clonal analysis of nodular parathyroid hyperplasia in renal hyperparathyroidism. World J Surg 1996;20:744–50; discussion 750–2.
5. Howell V, Haven C, Kahnoski K, et al. HRPT2 mutations are associated with malignancy in sporadic parathyroid tumors. J Med Genet 2003;40:657–63.
6. Carpten JD, Robbins C, Villablanca A, et al. HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. Nat Genet 2002;32:676–80.
7. Yano S, Sugimoto T, Tsukamoto T, et al. Decrease in vitamin D receptor and calcium-sensing receptor in highly proliferative parathyroid adenomas. Eur J Endocrinol 2003;148:403–11.
8. Solcia E, Kloppel G, Sobin LH. Histological Typing of Endocrine Tumours; International Histological Classification of Tumours. WHO Ed. 2. Springer; 2000. p. 48–55.
9. Haddad R, Furge KA, Miller J, et al. Genomic profiling and cDNA microarray analysis of human colon adenocarcinoma and associated intraperitoneal metastases reveals consistent cytogenetic and transcriptional aberrations associated with progression of multiple metastases. Appl Genom Proteom 2002;1:51–62.
10. Takahashi M, Yang XJ, Lavery TT, et al. Gene expression profiling of favorable histology Wilms tumors and its correlation with clinical features. Cancer Res 2002; 15:62:6598–605.
11. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8.
12. Rhodes DR, Miller JC, Haab BB, Furge KA. CIT: identification of differentially expressed clusters of genes from microarray data. Bioinformatics 2002;18:205–6.
13. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;24:98:5116–21.
14. Eilers PHC, Boer JM, Ommen van GJ, Houwelingen van HC. Classification of microarray data with penalized logistic regression. Proc SPIE; 2001;4266:187–98
15. Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol 2002;25:3:RESEARCH0036.
16. Wilson DL, Buckley MJ, Helliwell CA, Wilson IW. New normalization methods for cDNA microarray data. Bioinformatics 2003;19:1325–32.
17. Kiiveri HT. A Bayesian approach to variable selection when the number of variables is very large. In: Goldstrein DR, editor. Science and Statistics: A Festschrift for Terry Speed: Lecture Notes - Monograph Series. Hayward, California: Institute of Mathematical Statistics, 2003. pp. 127–43.
18. Bovee JV, van den Broek LJ, Cleton-Jansen AM, Hogendoorn PC. Up-regulation of PTHrP and Bcl-2 expression characterizes the progression ofosteochondroma towards peripheral chondrosarcoma and is a late event in centralchondrosarcoma. Lab Investig 2000;80:1925–34.
19. Kytola S, Farnebo F, Obara T, et al. Patterns of chromosomal imbalances in parathyroid carcinomas. Am J Pathol 2000;157:579–86.
20. Heppner C, Kester MB, Agarwal SK, et al. Somatic mutation of the MEN1 gene in parathyroid tumours. Nat Genet 1997;16:375–8.
21. Vasef MA, Brynes RK, Sturm M, Bromley C, Robinson RA. Expression of cyclin D1 in parathyroid carcinomas, adenomas, and hyperplasias: a paraffin immunohistochemical study. Mod Pathol 1999;12:412–6.
22. Downs JA, Kosmidou E, Morgan A, Jackson SP. Suppression of homologous recombination by the Saccharomyces cerevisiae linkerhistone. Mol Cell 2003;11:1685–92.
23. Pietrzik CU, Hoffmann J, Stober K, et al. From differentiation to proliferation: the secretory amyloid precursor proteinas a local mediator of growth in thyroid epithelial cells. Proc Natl Acad Sci USA 1998;17:95:1770–5.
24. Babich M, Foti LR. E-cadherins identified in osteoblastic cells: effects of parathyroid hormone and extracellular calcium on localization. Life Sci 1994;54:PL201–8.
25. Brabant G, Hoang-Vu C, Cetin Y, et al. E-cadherin: a differentiation marker in thyroid malignancies. Cancer Res 1993;5:53:4987–93.