

ORIGINAL ARTICLE

**Intracystic papillary carcinoma of breast harbors significant genomic alteration compared with intracystic papilloma: Genome-wide copy number and LOH analysis using high-density single-nucleotide polymorphism microarrays**

Masahiro Oikawa, MD,<sup>1,2</sup> Takeshi Nagayasu, MD, PhD,<sup>2</sup> Hiroshi Yano, MD, PhD,<sup>2</sup> Tomayoshi Hayashi, MD, PhD,<sup>3</sup> Kuniko Abe, MD, PhD,<sup>3</sup> Akira Kinoshita, PhD,<sup>1</sup> Koh-ichiro Yoshiura, MD, PhD\*<sup>1</sup>

*<sup>1</sup>Departments of Human Genetics and <sup>2</sup>Surgical Oncology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; <sup>3</sup>Department of Pathology, Nagasaki University Hospital, Nagasaki, Japan.*

*Address for reprints and correspondence:*

Koh-ichiro Yoshiura, Department of Human Genetics, Nagasaki University Graduate School of Biomedical Science, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Tel.: +81-95-819-7120

Fax: +81-95-819-7121

E-mail: kyoshi@nagasaki-u.ac.jp

**Running title**

*Molecular-cytogenetic Profile of ICPTs*

**KEY WORDS**

*Intracystic papillary tumor, breast cancer, array CGH, GeneChip®, FFPE*

## ABSTRACT

**Purpose:** Intracystic papillary breast tumors consist of benign papilloma, carcinoma in situ and carcinoma with invasion. Using high-density single-nucleotide polymorphism arrays, this study aimed to determine the profile of genomic alterations in these lesions and to identify novel diagnostic criteria. **Methods:** Ten samples of intracystic papillary tumor, which included five papillomas (Pap), three papillary carcinomas in situ (PurePC) and two papillary carcinomas with invasion (PCinv), were studied. DNA was extracted from tumor and normal tissues that were microdissected from the same formalin-fixed paraffin embedded blocks. Using probe intensity and genotype data from high-density oligonucleotide SNP microarrays (Affymetrix® GeneChip Genome-wide Human 5.0), paired copy number and LOH analysis was performed using Partek Genomic Suite Software. **Results:** Quality control (QC) call rate, which is an index measuring the quality of a SNP microarray experiment, ranged from 70.75% to 91.93%, mean 80.72%. The mean total genomic alteration rate (sum of amplifications, deletions and copy-neutral loss of heterogeneity) with respect to the whole genome was 2.87%, 15.4% and 35.3% in Pap, PC and IDC, respectively, and was significantly different between samples (Kruskal-Wallis chi-squared test,  $p = 0.043$ ). The most commonly altered regions ( $\geq 4/5$ ) in papillary carcinoma were copy-neutral loss of heterogeneity at 3p21.31 and 3p14.2 and amplification at 20q13.13. Genes altered only in invasive carcinoma included genes concerned with transcription. **Conclusions:** Among intracystic papillary breast tumors, malignant tumors, including non-invasive tumors, which are difficult to diagnose histopathologically, harbor significant genomic alteration. Our findings may aid clinical management of these tumors and may provide insight into their carcinogenesis.

## INTRODUCTION

Intracystic papillary breast tumors (ICPT) consist of benign papillomas, carcinomas in situ and carcinomas with invasion, and they account for approximately 10% of benign breast tumors and less than 1% of malignant tumors, respectively (1, 2). Intracystic papillary carcinoma develops predominantly in elderly women, who often present with a palpable mass and/or bloody nipple discharge. Although this type of carcinoma has a good prognosis, regardless of whether the tumor is diagnosed as in situ or invasive (3), some cases with metastasis to lymph nodes or to distant organs have been reported (4-6).

Cytogenetic studies of breast papillary tumors, which are considered useful for clinical diagnosis and for understanding tumorigenesis, are limited and cytogenetic differences between papillomas and papillary carcinomas are still controversial. Tsuda et al. (7, 8) reported that papillary carcinomas have frequent changes in gene copy-number and loss of heterozygosity (LOH), while papillomas did not show any gene copy-number alteration or LOH at 16q and 1q. Boecker et al. (9) also reported that conventional comparative genomic hybridization (CGH) did not reveal any gene copy-number change in papillomas. On the other hand, Lininger et al. (10) and Cristofano et al. (2) demonstrated that LOH at 16p or 16q was frequent in both papillomas and papillary carcinomas.

Recent technological progress has enabled CGH analysis with higher resolution using high-density array CGH (aCGH) (11). In breast cancer, aCGH studies have revealed genomic regions where DNA copy number is commonly changed. Novel candidate oncogenes or anti-oncogenes have been found in these regions, and relationships between genomic alteration and the clinical phenotypes and/or prognosis have been suggested (12-15). Analysis with high-density single nucleotide polymorphism (SNP) microarrays has the advantage of performing comprehensive genome-wide analyses of genomic alteration. The use of intensity data and genotype data from SNP-specific probes enables not only copy number detection but also copy number neutral LOH detection (16-19). Although this analysis requires high quality genomic DNA, such as DNA extracted from peripheral blood or rapidly-frozen samples, some modifications to the protocol and statistical processing have enabled us to use degraded genomic DNA extracted from formalin-fixed paraffin-embedded (FFPE)

samples (20-23). This enables the utilization of a large and growing deposit of archived clinical tissues that are stored as FFPE samples.

In this article, we reveal the profile of genomic alteration in breast ICPT using FFPE samples and show, for the first time, the possibility of using high-density oligonucleotide SNP arrays as the basis of a novel diagnostic method of ICPT.

## **MATERIALS AND METHODS**

### ***Tumor samples and clinical characteristics***

Ten FFPE breast ICPTs were obtained from the Department of Pathology, Nagasaki University Hospital. The samples included five benign papillomas (Pap), three papillary carcinomas in situ (PurePC) and two papillary carcinomas with invasion (PCinv). Pathological diagnosis was independently determined by two pathologists. Categorical diagnosis was determined from imaging studies by radiologists specializing in breast disease in accordance with the Breast Imaging-Reporting and Data System (BI-RADS) of the American College of Radiology and the diagnostic guidelines of the Japanese Association of Breast and Thyroid Sonology (JABTS). Clinicopathological findings of these tumors are provided in Figure 1, Table 1 and Supplementary Table 1.

All experimental procedures for this study were approved by the Committee for Ethical Issues on the Human Genome and Gene Analysis of Nagasaki University and all patients provided informed consent for voluntary participation.

### ***DNA extraction and hybridization to SNP arrays***

Using ten to twenty 10  $\mu\text{m}$  thick sections cut from FFPE blocks, tumor tissue areas containing more than 90% tumor cells and normal tissue areas not having any cancer cells, which were identified by staining with hematoxylin and eosin, were microdissected. Paraffin removal was performed in 80% xylene and samples were then washed twice with absolute ethanol. After drying the pellet was resuspended in 360  $\mu\text{L}$  of buffer ATL (QIAmp DNA Mini Kit, Qiagen, Germany) and incubated at 95°C for 15 minutes and then cooled to room temperature. Samples were then digested with proteinase K for 3 days at 56°C in a rotation oven with periodic mixing and the addition of fresh proteinase K every 24 hours.

DNA was collected using the QIAmp DNA Mini Kit according to the manufacturer's instructions. Briefly, 400  $\mu\text{L}$  of buffer AL was added to the sample and incubated at 70°C for 10 minutes. 400  $\mu\text{L}$  of absolute ethanol was then added. The sample solution was then placed into the spin column and centrifuged for 1 minute at 8000 x g. The spin column was washed twice with 500  $\mu\text{L}$  of AW1 by centrifugation at 8000 x g for one minute and then washed with AW2 by centrifugation at 14,000 x g for three minutes.

The DNA was finally eluted with 55  $\mu$ L buffer AE. Extracted DNA was quantified on a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). All samples used in this study had an OD 260/280 ratio higher than 1.8.

Extracted DNA from each sample was processed following the manufacturer's protocol and hybridized on Affymetrix GeneChip Genome-Wide Human SNP Array 5.0® (Affymetrix, Santa Clara, CA, USA). Because DNA extracted from FFPE samples was degenerated, the following modifications were adopted to the oligonucleotide microarray system, taking into consideration previous studies (20, 21, 23): the initial DNA amount was increased from 250 ng to 1  $\mu$ g; digestion time was prolonged from 120 minutes to overnight; the volume of PCR reactions was increased when the yields of PCR product failed to reach prescribed levels. The peak size of mapping PCR products was determined by visual inspection of electropherograms following 2% TBE agarose gel electrophoresis.

### *SNP array data analyses*

All signal intensities of probes and genotype calls were generated and obtained from Genotyping Console 3.0.1® (BRLMM-P algorithm) using default parameter settings. Overall hybridization quality was estimated by a generated QC call rate index. We use the term “copy number change” meaning deletion or amplification of a genomic region and “CNLOH” meaning copy number neutral loss of heterozygosity.

Copy number change and LOH analyses (called here SNPacGH) were conducted using the Partek Genomics Suite (PGS) version 6.3 (Partek, St. Louis, MI, USA). When signal intensities of probes were imported from CEL files into PGS, a normalization procedure with correction for GC-content and fragment length effects was performed. Copy number estimates from signal intensities were determined by “paired analysis”, to compare copy number state from tumor and matching normal tissue. Detection of amplifications and deletions was performed with a segmentation algorithm in the copy number workflow in PGS, where the minimum marker size was set at 150 (default setting is 10) and signal/noise ratio was set at 0.25 (default setting is 0.30). Genotype calls data were imported from CHP files into PGS following the restriction of SNPs to fragment sizes  $\leq$ 500 bp, because genotype calls of SNPs on longer fragments are

unreliable from degenerated DNA, such as DNA extracted from FFPE samples (20). LOH values were inferred by paired analysis with the Hidden Markov Model default setting in the LOH Workflow in PGS. All of these modifications from the default settings are adopted to maximize detection specificity.

### ***Quantitative PCR (qPCR) assay***

Quantitative PCR analyses to validate copy number changes were performed on a LightCycler® 480 Real-Time PCR System (Roche Diagnostics, Mannheim, Germany) using an intercalating dye, SYTO13 (Molecular probes, OR, USA), which is an alternative to SYBR green I. Absolute quantification was carried out using a second derivative max method (24). A standard calibration curve was generated with a serial dilution of genomic DNA to estimate the copy number state of sample for each set of primers. A corrected copy number state was given as the ratio of a target gene divided by an internal control gene. Copy number changes in tumors were determined by comparing paired samples (paired analysis).

Target genes for copy number validation and sequences of primer sets were as follows: ATP-binding cassette, sub-family A, member 5 (ABCA5, Forward; 5' TGCTGTGGTTCCCATCAAAC3' Reverse; 5' CATGCCAACACTCGTTCACA3'), G protein-coupled receptor 4 (GPR4, Forward; 5'AGGTGCAGCTGAAGATGCTG3' Reverse; 5'CTGTGGGATGAGAGGGGAAA3'), Frizzled 9 (FZD9, Forward; 5' TGCCCCCTCTCTGGCTACCTG3' Reverse; 5' GGGCACCGTG TAGAGGATGG3'), Snail 1 homolog (SNAI1, Forward; 5'CTAACCAGCTTGGAGGTGGG3' Reverse; 5'AGGGAGGACGTGACTGGTG3'). The diploid internal control gene and primer set sequences were Ornithine decarboxylase antizyme 2 (OAZ2, Forward; 5'CCTTCAGCTTCTTGGGCTTT3' Reverse; 5'TGGTCCAGGGGATAAACCAT3'). BLAST searches confirmed all primer sequences to be specific for the gene.

Samples were analyzed in quadruplicate in a 384-well format in a 10 µL final volume containing about 2 ng genomic DNA, 0.5 µM forward primer, 0.5 µM reverse primer, 0.1 Units TaKaRa ExTaq HS version (TaKaRa, Kyoto, Japan), 1 x PCR buffer, 200 µM dNTP and 0.5 µM SYTO13. The amplification conditions consisted of an initial denaturation at 95°C for 5 minutes, followed by 45 cycles of denaturation at 95°C

for 10 seconds, annealing at 55°C for 10 seconds and extension at 72°C for 15 seconds. The data were analyzed using LightCycler® 480 Basic Software (Roche Diagnostics) and melting curve analysis was always performed to verify the absence of non-specific amplification.

### ***Statistical analysis***

To estimate the total rate of a copy number changed region, each segment amplified or lost was summed and divided by 2,829 Mb, which is the total Mb in the genome, excluding heterochromatic, centromeric and telomeric regions not covered by probes. Similarly, to estimate the total rate of genomic alteration, the sum of segments with copy number change and CNLOH was divided by 2,829 Mb. Wilcoxon's rank sum test and Kruskal-Wallis' chi-squared test were performed to compare the rate of copy number change and genomic alteration between subgroups.

To determine successful predictive factors for the analysis of FFPE samples, Pearson's product-moment coefficient of correlation test was performed with the QC call rate.

The analyses above were done with the free statistical program, R (version 2.8.0) (<http://www.r-project.org/>) and the results were considered statistically significant when the p-value was <0.05. To determine what kinds of genes were contained in the gene lists, Gene Set Enrichment Analysis (GSEA) was performed with the H-InvDB Enrichment Analysis Tool (HEAT) (25). These results were considered statistically significant when the p-value for Fisher's exact probability test was <0.001.



## RESULTS

### *Performance of array hybridization*

The QC call rates obtained from the FFPE samples were from 70.75 to 91.93 %, with a mean of 80.72 % (Table 1, Supplementary Table 1). A highly significant linear correlation between the peak size of the mapping PCR product and the QC call rate was observed ( $r = 0.85$ ,  $P \leq 0.0001$ ). Also significant negative linear correlation between the duration of storage and the QC call rate was observed ( $r = -0.70$ ,  $P < 0.006$ ) (Supplementary figure 1). Yields of genomic DNA and PCR product showed no significant correlations with QC call rate (data not shown).

### *Genomic alterations detected by GeneChip Genome-Wide Human SNP 6.0 arrays and correlations with clinical characteristics*

In genome-wide copy number and LOH analysis, substantial divergence was observed between each ICPT subtype (Figure 2). The mean rate of copy number change was 0.48% (from 0.0% to 1.60%), 7.89% (from 0.41% to 12.0%) and 16.3% (from 16.0% to 16.6%) in Pap, PC and PCinv, respectively. The mean rate of genomic alteration (including copy number change and CNLOH) was 2.87% (from 0.00% to 11.8%), 15.4% (from 8.83% to 24.1%) and 35.3% (from 17.6% to 53.1%) in Pap, PC and PCinv, respectively (Table 1). Malignant tumors (PurePC and PCinv) showed significantly more copy number changes and genomic alterations (copy number change and CNLOH) than benign tumors (Pap) (Wilcoxon's rank sum test,  $p = 0.036$ ,  $0.016$  respectively) (Figure 3) and these differences correlated with their malignant phenotype (Kruskal-Wallis' chi-squared test,  $p = 0.046$ ,  $0.043$ , respectively) (Fig 4).

### *Copy number validation by qPCR assay*

To validate the copy number change identified by SNPacGH, quantitative PCR assays were performed at four selected loci, including independent genes (Table 2). All loci showing alteration were confirmed by real-time qPCR. This indicated that the detection specificity of copy number change in SNPacGH was 100%. On the other hand, at 31 loci from the ten samples, where SNPacGH showed the copy number state as disomy, ten loci were revealed to have a copy number change by real-time qPCR. At regions determined to be disomy by SNPacGH, 70.6 % were determined by qPCR to be

two copy. The copy number calling concordant rate between benign (Pap) and malignant (PurePC + PCinv) tumors was not significant (Fisher's exact test,  $p = 0.31$ ), at 73.7% and 58.3% respectively.

### ***Genes within frequently altered chromosomal regions***

Among five carcinomas we identified 3 or more that had 93 regions of chromosomal alteration (Supplementary table 2), ranging in size from 0.7 kb to 4.8 Mb (median 215 kb), which involved 641 RefSeq genes. Chromosomal regions at 3p21.31, 3p14.2 and 20q13.13 were commonly altered ( $\geq 4/5$ ) and 18 RefSeq genes were involved (Table 3). The 326 RefSeq genes involved only in PCinv are listed in Supplementary table 3 as “invasion genes list” because they are genes responsible for cancer invasion. Data analysis by GSEA revealed that this list included genes concerned with transcription, such as nucleotide binding (GO ID 0000166), cell communication (GO ID 0007154) and ATP binding (GO ID 0005524) (Table 4).

## DISCUSSION

In breast lesions, indication for surgery is usually determined by pathological diagnosis together with radiological findings but differential, preoperative diagnosis of papillary carcinoma from papilloma is very difficult, even following needle biopsy (26) because of their non-specific radiological characteristics and their modest cytological and histological appearance (6). Hence, in the clinical management of these lesions, surgical excision is recommended, especially when associated to identified risk factors of malignancy, such as high age ( $\geq 50$  years) and the presence of microcalcifications (27, 28). In the cases presented in this study, we also needed to conduct surgery since we couldn't determine lesion malignancy (Figure 1, Table 1). To avoid excessive surgical intervention, another diagnostic procedure needs to be developed.

In the present study, we have compared the molecular profiles of papilloma versus papillary carcinoma through a genome-wide copy number and LOH analysis using new technology: high-density oligonucleotide SNP microarrays. Although several studies have compared the molecular profiles of papilloma versus papillary carcinoma (2, 7-10), the differences between these profiles were not conclusive. Some studies indicated that papillary carcinoma harbored more genomic alterations than papilloma (7-9) but other studies indicated that papilloma also had substantial genomic alterations (2, 10). Our SNP aCGH results indicated that papillary carcinoma harbored significantly more genomic alterations than papilloma, even though papilloma had a number of genomic alterations, and that the rate of genomic alteration correlated with pathological malignancy classification. Our findings also suggest that SNP aCGH could be a new preoperative diagnostic method for papillary lesions. We need to analyze more samples to confirm our findings. Also a prospective study using high-density SNP arrays and specimens from preoperative core-needle biopsies is required for practical clinical application.

Previous studies have documented that the most common genomic alteration in papillary carcinoma is amplification on 1p and deletion or LOH on 16q (2, 7, 8, 10). Our study revealed deletion or CNLOH on 3p and amplification on 20q in addition to deletion or CNLOH on 16q (Figure 2, Table 3, Supplementary table 2). Intriguingly, this profile is similar to intraductal or invasive ductal carcinoma(9) and these regions

contain *SNAIL* previously linked to breast cancer biology(29). The significance of 3p and 20q are currently unclear but require further investigation.

Some interesting genes are included in the “invasion gene list”. MAML1 (30), located at 5q35, is a transcriptional coactivator for NOTCH proteins, E2F1 (31), located at 20q11.22, is a transcriptional activator that cooperatively binds DNA with dp proteins through the E2 recognition site and TPT1 (32), located at 13q14.12, is involved in calcium binding and microtubule stabilization. These genes have been previously reported to be involved in cancer. Thus, this list identifies genes for further mechanistic research and could include several future therapeutic targets.

In this study, we applied degraded DNA, extracted from FFPE samples, to high-density SNP arrays. When degraded DNA is used with the Affymetrix GeneChip system the intensities of probe signals hybridized to long fragments tend to be weakened, which would cause artificial copy number change. To resolve this problem, previous studies modified the extraction protocol and/or filtered out signals expected to result from hybridization of long DNA fragments (20-23, 33). Referring to these studies, we have adopted the modifications described above. The QC call rate obtained in this study ranged from 70.75 to 91.93%, with a mean of 80.72%, which was comparable to the results from former cytogenetic studies using DNA extracted from FFPE samples. For “paired analysis”, QC call rate would not be important for copy number state estimation and LOH detection. QC call rate itself is calculated from the call of SNPs that are relatively difficult to genotype, even using high quality DNA. Because concordance between tumor and non-tumor SNP call is important to detect LOH with or without deletion, a critical factor using FFPE samples is to use tumor and non-tumor tissues from the same slice samples. “Paired analyses” from the same FFPE sample will assure the concordance of SNP calls and hybridization intensities. Moreover, validation of the copy number state in SNPacGH by real time qPCR has demonstrated good specificity. The relatively low concordant copy number state between SNPacGH and qPCR in disomy regions could be caused by the copy number detection setting, which was adopted to minimize false positive detection and maximize detection specificity. For example, a tiny region with copy number change that is detected by real time qPCR would be overlooked by SNPacGH analysis using this setting. Even if this experimental error is taken into consideration, our results of

significant genomic alteration difference between papilloma and papillary carcinoma is still acceptable because the sensitivity in papilloma and in papillary carcinoma is not significantly different. Thus FFPE samples, which are accessible and associated with much clinicopathological information, will be useful, especially for cytogenetic analyses of rare phenotypes. In addition, the methodology described here can be adapted to other kinds of carcinoma.

In summary, we have elucidated significant differences in the molecular-cytogenetic profile between papilloma and papillary carcinoma from FFPE samples by genome-wide copy number and LOH analysis using high-density single-nucleotide polymorphism microarrays. Our data encourage us to exploit a vast number of archived FFPE samples to investigate the biology of a variety of cancers, including breast cancer. The genes contained in common altered regions are fascinating candidates for further research to unravel the mechanisms of tumorigenesis and invasion of breast cancer.

#### **ACKNOWLEDGMENTS**

We are grateful to the patients for their participation in this research. We also thank Ms M. Ooga and Ms C. Hayashida for their technical assistance. K.Y. was supported partly by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labor and Welfare, from the Ministry of Education, Culture, Sports, Science and Technology of Japan Category B, No. 21390100), and by grants from the Takeda Scientific Foundation and the Naito Foundation.

#### **COMPETING INTERESTS**

There are no competing interests

#### **PATIENT CONSENT**

Obtained

## REFERENCES

1. Fayanju OM, Ritter J, Gillanders WE, *et al.* Therapeutic management of intracystic papillary carcinoma of the breast: the roles of radiation and endocrine therapy. *Am J Surg* 2007;194:497-500.
2. Di Cristofano C, Mrad K, Zavaglia K, *et al.* Papillary lesions of the breast: a molecular progression? *Breast Cancer Res Treat* 2005;90:71-76.
3. Grabowski J, Salzstein SL, Sadler GR, Blair S. Intracystic papillary carcinoma: a review of 917 cases. *Cancer* 2008;113:916-920.
4. Okita R, Ohsumi S, Takashima S, *et al.* Synchronous liver metastases of intracystic papillary carcinoma with invasion of the breast. *Breast Cancer* 2005;12:327-330.
5. Mulligan AM, O'Malley FP. Metastatic potential of encapsulated (intracystic) papillary carcinoma of the breast: a report of 2 cases with axillary lymph node micrometastases. *Int J Surg Pathol* 2007;15:143-147.
6. Solorzano CC, Middleton LP, Hunt KK, *et al.* Treatment and outcome of patients with intracystic papillary carcinoma of the breast. *Am J Surg* 2002;184:364-368.
7. Tsuda H, Takarabe T, Susumu N, *et al.* Detection of numerical and structural alterations and fusion of chromosomes 16 and 1 in low-grade papillary breast carcinoma by fluorescence in situ hybridization. *Am J Pathol* 1997;151:1027-1034.
8. Tsuda H, Takarabe T, Akashi-Tanaka S, Fukutomi T, Hirohashi S. Pattern of chromosome 16q loss differs between an atypical proliferative lesion and an intraductal or invasive ductal carcinoma occurring subsequently in the same area of the breast. *Mod Pathol* 2001;14:382-388.
9. Boecker W, Buerger H, Schmitz K, *et al.* Ductal epithelial proliferations of the breast: a biological continuum? Comparative genomic hybridization and high-molecular-weight cytokeratin expression patterns. *J Pathol* 2001;195:415-421.
10. Lininger RA, Park WS, Man YG, *et al.* LOH at 16p13 is a novel chromosomal alteration detected in benign and malignant microdissected papillary neoplasms of the breast. *Hum Pathol* 1998;29:1113-1118.
11. Climent J, Garcia JL, Mao JH, Arsuaga J, Perez-Losada J. Characterization of breast cancer by array comparative genomic hybridization. *Biochem Cell Biol* 2007;85:497-508.
12. Naylor TL, Greshock J, Wang Y, *et al.* High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Res* 2005;7:R1186-1198.
13. Fridlyand J, Snijders AM, Ylstra B, *et al.* Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006;6:96.
14. Andre F, Job B, Dessen P, *et al.* Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. *Clin Cancer Res* 2009;15:441-451.
15. Reis-Filho JS, Simpson PT, Gale T, Lakhani SR. The molecular genetics of breast cancer: the contribution of comparative genomic hybridization. *Pathol Res Pract* 2005;201:713-725.

16. Yamamoto G, Nannya Y, Kato M, *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* 2007;81:114-126.
17. Zhao X, Li C, Paez JG, Chin K, *et al.* An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004;64:3060-3071.
18. Wang ZC, Buraimoh A, Iglehart JD, Richardson AL. Genome-wide analysis for loss of heterozygosity in primary and recurrent phyllodes tumor and fibroadenoma of breast using single nucleotide polymorphism arrays. *Breast Cancer Res Treat* 2006;97:301-309.
19. Argos M, Kibriya MG, Jasmine F, *et al.* Genomewide scan for loss of heterozygosity and chromosomal amplification in breast carcinoma using single-nucleotide polymorphism arrays. *Cancer Genet Cytogenet* 2008;182:69-74.
20. Jacobs S, Thompson ER, Nannya Y, *et al.* Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. *Cancer Res* 2007;67:2544-2551.
21. Lyons-Weiler M, Hagenkord J, Sciulli C, Dhir R, Monzon FA. Optimization of the Affymetrix GeneChip Mapping 10K 2.0 Assay for routine clinical use on formalin-fixed paraffin-embedded tissues. *Diagn Mol Pathol* 2008;17:3-13.
22. Nessling M, Richter K, Schwaenen C, *et al.* Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res* 2005;65:439-447.
23. Thompson ER, Herbert SC, Forrest SM, Campbell IG. Whole genome SNP arrays using DNA derived from formalin-fixed, paraffin-embedded ovarian tumor tissue. *Hum Mutat* 2005;26:384-389.
24. Luu-The V, Paquet N, Calvo E, Cumps J. Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. *Biotechniques* 2005;38:287-293.
25. Yamasaki C, Murakami K, Takeda J, *et al.* H-InvDB in 2009, extended database and data mining resources for human genes and transcripts. *Nucleic Acids Research* 2009;38 (In press.)
26. Douglas-Jones AG, Verghese A. Diagnostic difficulty arising from displaced epithelium after core biopsy in intracystic papillary lesions of the breast. *J Clin Pathol* 2002;55:780-783.
27. Sakr R, Rouzier R, Salem C, *et al.* Risk of breast cancer associated with papilloma. *Eur J Surg Oncol* 2008;34:1304-1308.
28. Rizzo M, Lund MJ, Oprea G, *et al.* Surgical follow-up and clinical presentation of 142 breast papillary lesions diagnosed by ultrasound-guided core-needle biopsy. *Ann Surg Oncol* 2008;15:1040-1047.
29. Olmeda D, Moreno-Bueno G, Flores JM, *et al.* SNAIL1 is required for tumor growth and lymph node metastasis of human breast carcinoma MDA-MB-231 cells. *Cancer Res* 2007;67(24):11721-11731.
30. Wu L, Aster JC, Blacklow SC, *et al.* MAML1, a human homologue of Drosophila mastermind, is a transcriptional co-activator for NOTCH receptors. *Nat Genet* 2000;26:484-489.

31. Worku D, Jouhra F, Jiang GW, *et al.* Evidence of a tumour suppressive function of E2F1 gene in human breast cancer. *Anticancer Res* 2008;28:2135-2139.
32. Tuynder M, Susini L, Prieur S, *et al.* Biological models and genes of tumor reversion: cellular reprogramming through tpt1/TCTP and SIAH-1. *Proc Natl Acad Sci USA* 2002;99:14976-14981.
33. Tuefferd M, De Bondt A, Van Den Wyngaert I, *et al.* Genome-wide copy number alterations detection in fresh frozen and matched FFPE samples using SNP 6.0 arrays. *Genes Chromosomes Cancer* 2008;47:957-964.



Table 1. Characteristics of ten intracystic papillary lesions

Case	Clinicopathological findings						Genetic findings			
	Diagnosis	age	Size of cyst (mm)	MMG	US	FNAC	Receptor status	QC call rate (normal/tumor)	ratio of CNC	ratio of genomic alteration
1	Pap	43	80	Category 3	Category 3	Class 2	ER(+)	75.9%/82.9%	0.14%	0.24%
2	Pap	38	10	Category 1	Category 3	Class 3	NaN	83.4%/80.4%	0.66%	0.69%
3	Pap	49	25	Category 3	Category 3	Class 3	NaN	86.2%/86.5%	1.60%	1.60%
4	Pap	38	70	Category 3	Category 3	Class 2	NaN	89.9%/87.9%	0%	11.8%
5	Pap	49	75	Category 3	Category 3	Class 2	NaN	91.9%/89.8%	0%	0%
6	PurePC	61	31	Category 4	Category 3	Class 4	ER(+), HER2(1+)	75.7%/76.2%	11.3%	24.1%
7	PurePC	58	49	Category 3	Category 4	Class 4	ER(+)	79.7%/70.8%	0.41%	8.83%
8	PurePC	43	16	Category 2	Category 4	Class 4	ER(+), HER2(1+)	77.2%/79.9%	12.0%	13.2%
9	PCinv	60	96	NaN	Category 4	Class 1	ER(-), HER2(1+)	71.6%/73.9%	16.6%	53.1%
10	PCinv	72	19	Category 4	Category 4	Class 5	ER(+), HER2(1+)	82.0%/72.6%	16.0%	17.6%

Pap: intracystic papilloma, Pure PC: intracystic papillary carcinoma in situ, PC inv: intracystic papillary carcinoma with invasion, MMG: the mammographic features evaluated according to the BI-RADS, US: the ultrasonographic features evaluated according to diagnostic guideline of JABTS, FNAC: the cytological features of fine needle aspiration cytology, ER: the status of estrogen receptor, HER2: the status of HER2/neu receptor, CNC: copy number change, genomic alteration: copy number change and copy neutral loss of heterozygosity, NaN: not analyzed.

Table 2. The validation of Array-based comparative genomic hybridization by real time quantitative PCR

Case	ABCA5			GPR4			FZD9			SNAI1		
	aCGH	Light Cycler		aCGH	Light Cycler		aCGH	Light Cycler		aCGH	Light Cycler	
	Copy number state	Relative ratio*	Relative alleles	Copy number state	Relative ratio*	Relative alleles	Copy number state	Relative ratio*	Relative alleles	Copy number state	Relative ratio*	Relative alleles
1	deletion	0.58±0.15	1	disomy	1.33±0.52	3	disomy	1.55±0.23	3	disomy	0.95±0.31	2
2	disomy	0.76±0.17	2	disomy	1.05±0.19	2	disomy	1.16±0.37	2	disomy	0.98±0.48	2
3	disomy	0.87±0.08	2	disomy	1.22±0.10	2	disomy	1.60±0.22	3	disomy	1.11±0.63	2
4	disomy	2.26±1.10	5	disomy	1.45±0.76	3	disomy	0.91±0.48	2	disomy	1.24±0.70	2
5	disomy	1.14±0.16	2	disomy	0.86±0.11	2	disomy	0.80±0.31	2	disomy	0.97±0.13	2
6	disomy	0.97±0.17	2	disomy	1.01±0.12	2	disomy	1.11±0.76	2	disomy	0.96±0.66	2
7	disomy	1.64±1.16	3	disomy	0.52±0.09	1	disomy	0.83±0.17	2	amplification	2.55±0.89	5
8	disomy	0.79±0.12	2	disomy	2.40±0.39	5	disomy	2.96±0.57	4	amplification	1.56±0.36	3
9	disomy	1.30±0.11	3	amplification	4.87±0.38	10	amplification	2.87±0.29	5	amplification	1.78±0.99	4
10	disomy	1.03±0.32	2	amplification	1.86±0.71	4	amplification	1.86±0.31	4	amplification	2.03±1.07	4

\*: relative ratio is indicated mean ± coefficient variance.

Table 3. Common altered ( $\geq 4/5$ ) genes in intracystic papillary carcinoma

Chromosome	Cyto band	Gene symbol	Refseq accession	Description	status
3	p21.31	<i>CACNA2D2</i>	NM_006030.2	calcium channel, voltage-dependent, alpha	Copy-Neutral LOH
3	p21.31	<i>C3orf18</i>	NM_016210	hypothetical protein LOC51161	Copy-Neutral LOH
3	p21.31	<i>HEMK1</i>	NM_016173	HemK methyltransferase family member 1	Copy-Neutral LOH
3	p21.31	<i>CISH</i>	NM_013324	cytokine-inducible SH2-containing protein	Copy-Neutral LOH
3	p21.31	<i>MAPKAPK3</i>	NM_004635	mitogen-activated protein kinase-activated	Copy-Neutral LOH
3	p21.31	<i>DOCK3</i>	NM_004947	dedicator of cytokinesis 3	Copy-Neutral LOH
3	p14.2	<i>FAM3D</i>	NM_138805	family with sequence similarity 3, member D	Copy-Neutral LOH
20	q13.13	<i>KCNB1</i>	NM_004975	potassium voltage-gated channel, Shab-related	Amplification
20	q13.13	<i>B4GALT5</i>	NM_004776	UDP-Gal:betaGlcNAc beta 1,4-	Amplification
20	q13.13	<i>SPATA2</i>	NM_006038	spermatogenesis associated 2	Amplification
20	q13.13	<i>UBE2V1</i>	NM_022442	ubiquitin-conjugating enzyme E2 variant 1	Amplification
20	q13.13	<i>TMEM189-UBE2V1</i>	NM_199203	TMEM189-UBE2V1 readthrough transcript isoform1	Amplification
20	q13.13	<i>TMEM189</i>	NM_199129	transmembrane protein 189	Amplification
20	q13.13	<i>SNAI1</i>	NM_005985	snail 1 homolog	Amplification
20	q13.13	<i>PTGIS</i>	NM_000961	prostaglandin I2 (prostacyclin) synthase	Amplification
20	q13.13	<i>RNF114</i>	NM_018683	zinc finger protein 313	Amplification
20	q13.13	<i>SLC9A8</i>	NM_015266	sodium/hydrogen exchanger 8	Amplification
20	q13.13	<i>CEBPB</i>	NM_005194	CCAAT/enhancer binding protein beta	Amplification

Table 4. The result of Gene Set Enrichment Analysis (GSEA)

GO ID*	Name	Gene Symbol	P-value
0000166	nucleotide binding	<i>RBM12, INTS6, ARL11, DMPK, RASL11A, ABCA6, STK10, PMS2L3, PABPC3, CPNE1, MYH7B, RBM39, CHD5, TRPC4AP, CDK8, TUBA3C, ATP7B, RAP2A, GSS, MYLK2, DOCK9, ACSS2, SEPT9, ABCA8, RFC3, RAC3, HCK, EIF4H, ATP8A2, LIMK1, KIF3B, ABCA9, THOC4, TPX2, RAGE, NCF1, C20orf112, ELN, MAP2K6, GTF2F2, DUS1L, RFC2</i>	0.000320
0007154	cell communication	<i>FREM2, NCF1, SH3PXD2B, SNX5</i>	0.000700
0005524	ATP binding	<i>INTS6, ATP6V0E1, CHD5, TRPC4AP, ATP7B, ABCA8, ATP8A2</i>	0.000799
0035091	phosphoinositide binding	<i>PIGU, SNX5, NCF1, SH3PXD2B,</i>	0.00100
0005737	cytoplasm	<i>ACOT7, ACSS2, ARHGDI, BNIP1, C1QTNF9, CANX, DSTN, ELN, EPB41L1, ERFF1, HGS, KCNAB2, KPN A3, LIMK1, NDRG3, PABPC3, PARK7, POFUT1, RAG E, RCBTB1, RUFY1, SEPT9, SLA2, SYMPK, TPP2, TPT1,</i>	0.00428
0005525	GTP binding	<i>TPX2, NCF1, C20orf112, ELN, ARL11, RASL11A, TUBA3C, RAP2A, DOCK9, SEPT9, RAC3</i>	0.00517
0017111	nucleoside-triphosphatase activity	<i>ABCA9, NCF1, RFC2, INTS6, ATP6V0E1, ABCA6, CHD5, TRPC4AP, TUBA3C, ATP7B, RAP2A, SEPT9, ABCA8, RFC3, RAC3, ATP8A2</i>	0.00524
0005242	inward rectifier potassium channel activity	<i>KCNJ16, KCNJ2</i>	0.00676
0007264	small GTPase mediated signal transduction	<i>ARL11, RAC3, RALGPS1, RAP2A, RASL11A, RIN2</i>	0.00716

\*: GO ID: the ID assigned in the Gene Ontology project (<http://www.geneontology.org/>)

Supplementary Table 1. Backgrounds of samples affected on QC call rate.

Case	Diagnosis	subtype	Storage time (month)	Yields of genomic DNA ( $\mu$ g)	Yields of PCR product ( $\mu$ g)	Peak size of PCR product (bp)	QC call rate (%)
1	Pap	tumor	33	26.3	280	370	75.94
		normal	33	7.28	267.2	390	82.86
2	Pap	tumor	23	5.56	310	390	83.42
		normal	23	1.93	307.5	390	80.41
3	Pap	tumor	6	24.1	203.6	489	86.2
		normal	6	5.02	273.3	489	86.5
4	Pap	tumor	15	2.30	191.1	489	89.91
		normal	15	3.39	208.3	489	87.89
5	Pap	tumor	11	7.12	462.3	711	91.93
		normal	11	13.2	433.7	711	89.84
6	PurePC	tumor	54	8.13	206.8	300	75.68
		normal	54	45.7	280.3	242	76.21
7	PurePC	tumor	49	1.51	288.8	404	79.68
		normal	49	75.2	301.3	300	70.75
8	PurePC	tumor	33	11.96	290.5	404	77.2
		normal	33	8.36	209.1	341	79.88
9	PCinv	tumor	30	15.83	210.2	300	71.64
		normal	30	32.2	222.9	300	73.86
10	PCinv	tumor	16	4.92	323.7	341	82.03
		normal	16	3.32	511.4	341	72.57

Pap intracystic papilloma, Pure PC intracystic papillary carcinoma in situ, PC inv intracystic papillary carcinoma with invasion

## FIGURE LEGENDS

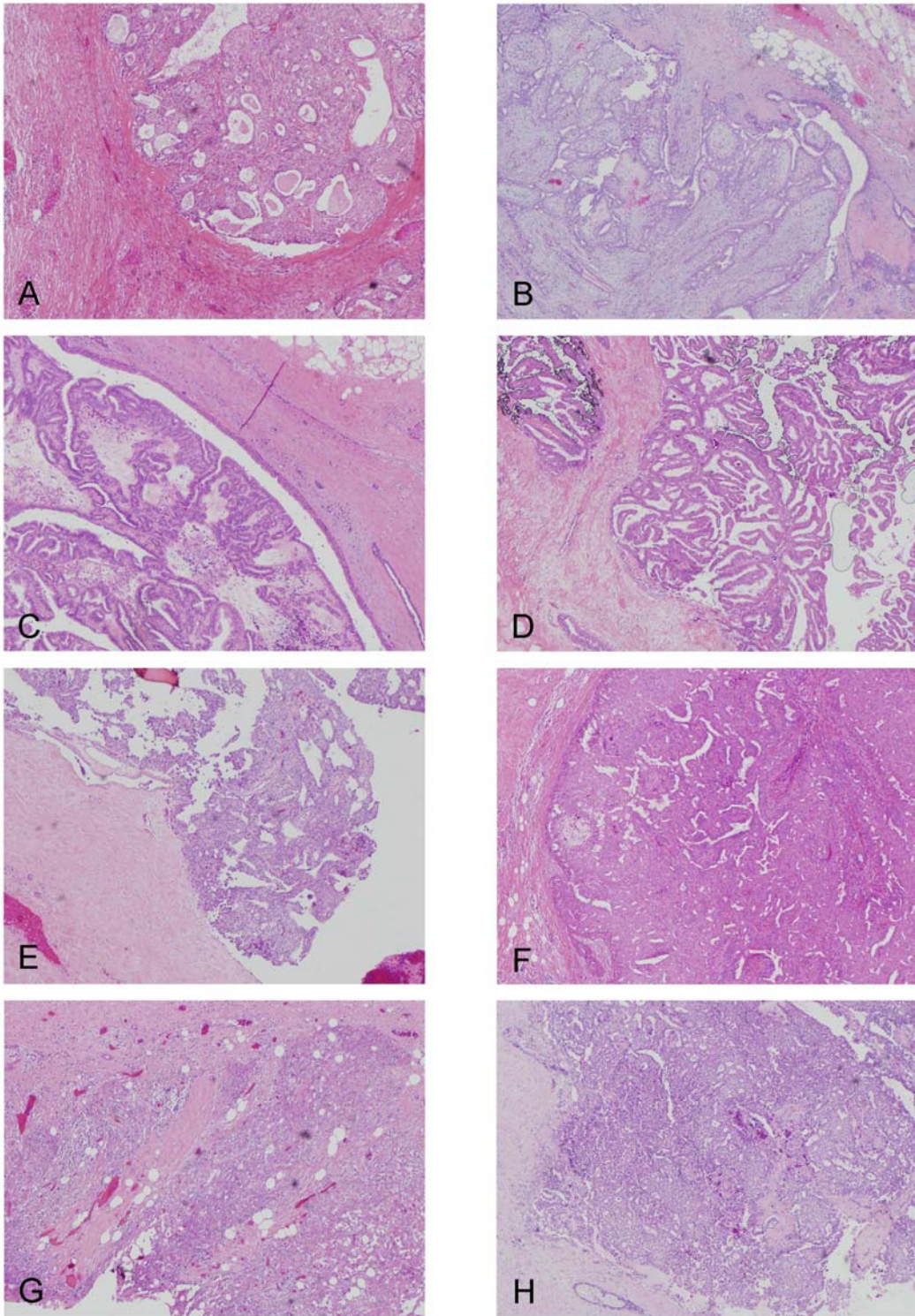


Figure 1. Hematoxylin-eosin stain in intracystic papillary tumors (Original magnification x40). A- C, Intracystic papilloma (A: case 1, B: case 2, C: case 3). D-F, Intracystic papillary carcinoma in situ (D: case 6, E: case 7, F: case 8). G and H, Intracystic papillary carcinoma with invasion (G: case 9, H: case10).



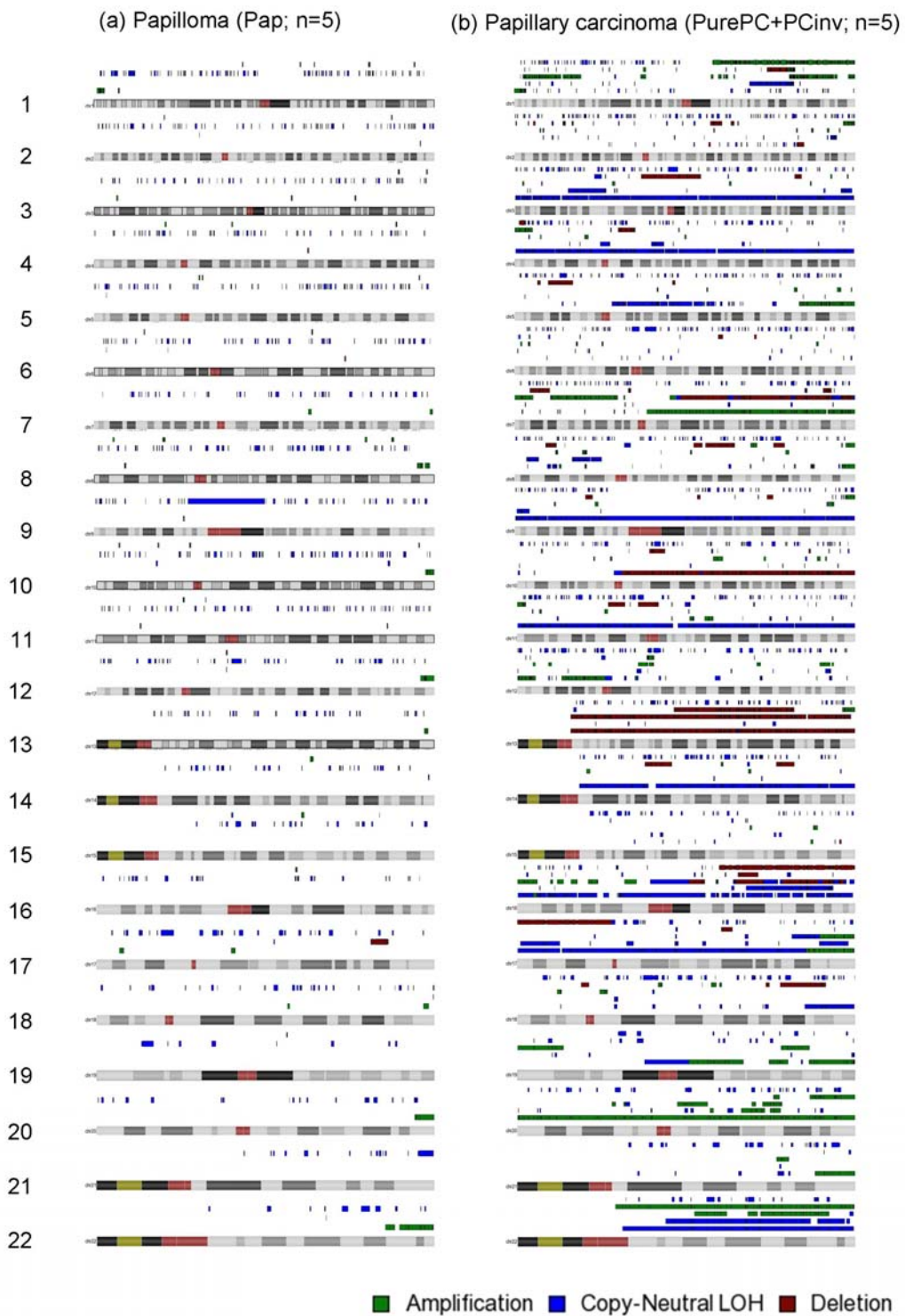


Figure 2. Graphic display of whole genomic alterations in papilloma (a) and papillary carcinoma (b). The color bar over each chromosome indicates copy number amplification (green color bars), copy-neutral LOH (blue color bars) and deletion (brown color bars) for each case. Papilloma includes 5 cases of Pap (a), and papillary carcinoma includes 3 cases of PurePC and two of PCInv (b).

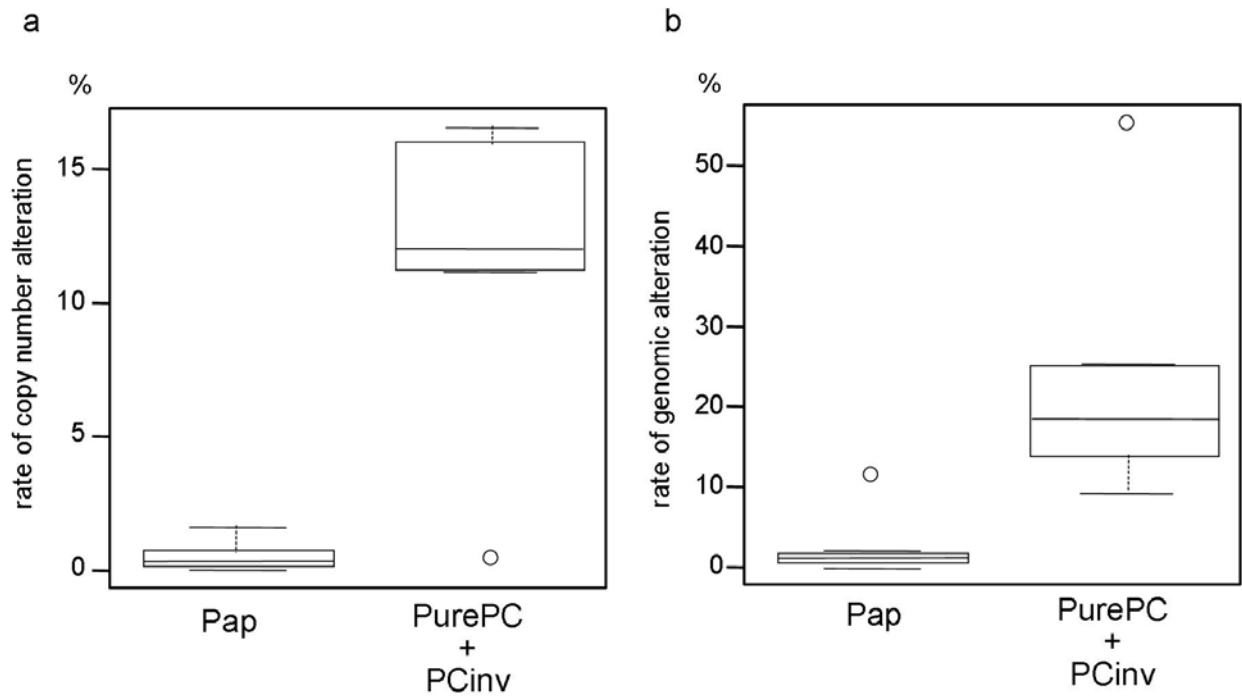


Figure 3. The mean rate of copy number change (a) and genomic alteration (b) in papilloma was compared with that in papillary carcinoma. The central rectangle spans the first quartile to the third quartile. A segment inside the rectangle shows the median and "whiskers" above and below the box show the locations of the minimum and maximum. The white circles represent outliers.



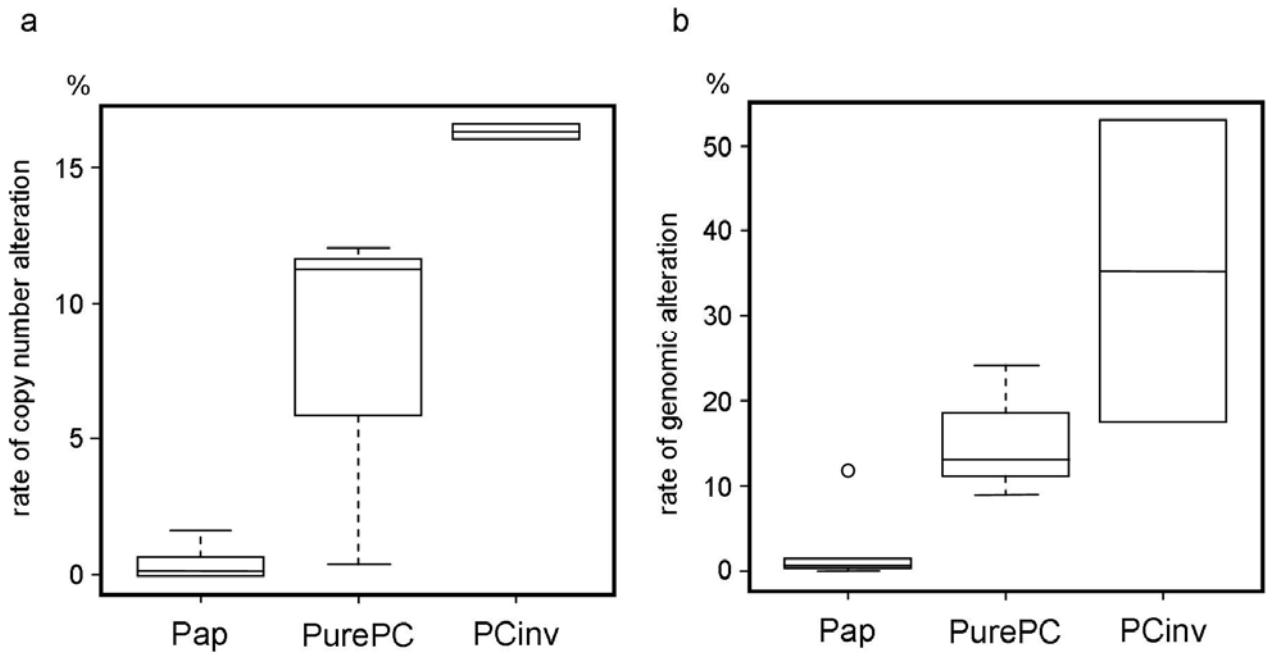
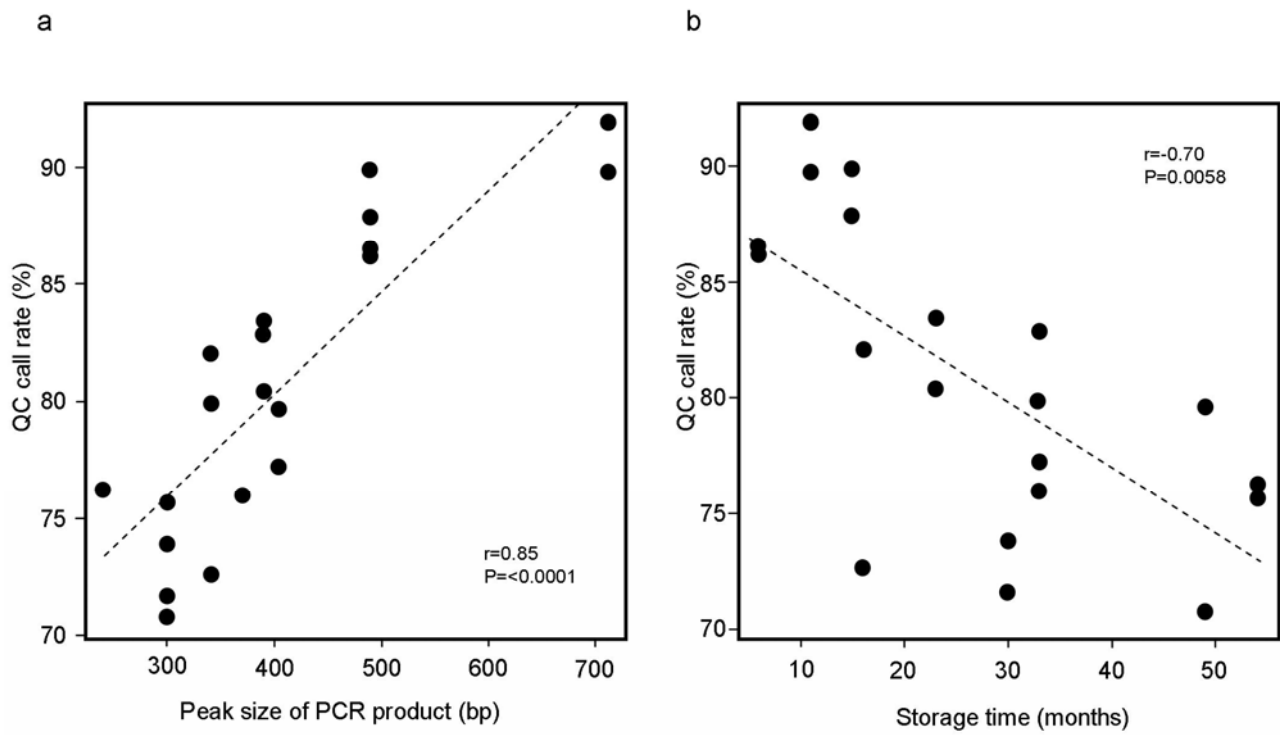


Figure 4. The comparison of mean rate of copy number change (a) and genomic alteration (b) among the three subtypes (Pap: papilloma, PurePC: papillary carcinoma in situ, PCinv: papillary carcinoma with invasion). The central rectangle spans the first quartile to the third quartile. A segment inside the rectangle shows the median and "whiskers" above and below the box show the locations of the minimum and maximum. The white circle over the rectangle represents an outlier.



Supplementary figure 1.

Correlation of QC call rate with peak size of PCR product (a) or storage time of FFPE blocks (b)