Research article

Copy number variations: A novel molecular marker for papillary thyroid cancer

Xingjian Lai [a,1], Luying Gao [a,1], Gaoying Zhou [b], Xiequn Xu [c,*], Jinhui Wang [d]

[a] Department of Ultrasound, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[b] Beijing Longer Gene Technology Co., Ltd., Beijing, China
[c] Department of General Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[d] Department of Gynecology and Obstetrics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

HIGHLIGHTS

1. Thyroid nodules can be identified by CNVs at the corresponding positions on chromosomes 5, 7, 8, 10, and 17.
2. The identification rate of PTC can be greatly increased through high-volume CNV sequencing analysis.

ARTICLE INFO

ABSTRACT

Background: We aimed to screen tumor-associated functional genes on a large scale through copy number variation (CNV) analysis of papillary thyroid cancer (PTC).
Methods: We analyzed 74 tissue samples from 41 patients with thyroid nodules. The samples were subjected to whole-genome resequencing and then analyzed by the 'WISECONDOR' method. Potential chromosome CNV regions were identified between the different sample groups.
Results: Of the 74 samples from 41 patients, 28 were PTC tissue samples, 29 were para-carcinoma tissue samples, 13 were benign tumor tissue samples and 4 were para-benign tumor tissue samples. According to our findings, PTC can be identified by CNVs at the corresponding positions on chromosomes 5, 7, 8, 10, and 17. For carcinoma tissue, the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and area under the curve (AUC) of the test method were 100%, 66.7%, 87.5%, 100.0%, 90.0% and 0.83 (95% confidence interval [CI], 0.67–1.00) and for para-carcinoma tissue, these values were 96.6%, 75.0%, 96.6%, 75.0%, 93.9% and 0.86 (95% CI, 0.60–1.00).
Conclusion: CNV analysis assays involving high-volume sequencing analysis can increase the identification of PTC, potentially avoiding errors caused by position deflection in sampling.

1. Introduction

Thyroid nodules are common, with a prevalence ranging from 19% to 68% in the general population. Approximately 7%–15% of thyroid nodules are thyroid cancer, and the incidence of papillary thyroid cancer (PTC), the most common type of thyroid cancer, has increased worldwide, accounting for 85% of differentiated thyroid cancers [1, 2]. Ultrasound (US)-guided fine-needle aspiration biopsy (FNAB) is the dominant and preferred screening modality. However, FNAB does not provide a differential diagnosis between malignant and benign lesions in approximately 25% of cases [2, 3]. The molecular characterization of cancers provides an opportunity to address this problem. Therefore, the screening of several potential molecular markers, such as BRAF, P53, and PIK3CA gene mutations, TERT promoter mutation, and NTRK1 gene recombination, is a fairly common practice [4, 5]. However, according to the American Thyroid Association (ATA) guidelines, the utility of molecular testing is applicable only when combined with the assessment of clinical and ultrasonic risk factors for malignancy [2].

Oligonucleotide arrays allow the detection of copy number variations (CNVs), a structural variation phenomenon that widely exists in the human genome, with high resolution on a genome-wide scale [6]. Recent studies of cancer genomics have found that distinct patterns in CNVs can be used to discriminate lung cancer, endometrial cancer, gastric cancer, and prostate cancer [7, 8, 9, 10]. This study focused on the CNV regions of tumor-related genes and found a more sensitive and specific molecular marker of PTC; our results may provide novel insight into the characteristics of PTC onset and possibly provide clues regarding the mechanisms underlying the risk for thyroid cancer.

2. Materials and methods

This study was approved by the ethics committee at our center. All procedures were conducted in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Written informed consent for participation was required for this study in accordance with national legislation and institutional requirements.

2.1. Study design

Forty-one northern Han Chinese with 74 available tissue samples who were long-term residents of Beijing were recruited at our center. The inclusion criteria were as follows: (i) total or near-total thyroidectomy and (ii) pathology-based diagnosis of thyroid tumor. Patients were excluded if they had undergone previous thyroid resection or treatment at another institution for PTC of any size or had an uncertain diagnosis.

2.2. DNA extraction

Genomic DNA was extracted from fresh tissue using a DNA extraction kit (TIANamp Genomic DNA Kit DP304) according to the manufacturer's instructions. Sample purity was assessed by a NanoPhotometer® (IMPLEN, CA, USA), and the DNA concentration was measured with a Qubit® 3.0 Fluorometer (Life Technologies, CA, USA). DNA fragments with a length of approximately 350 bp were randomly produced by Covaris ultrasonic DNA shearing.

2.3. Quality control of library and sequencing

The Agilent SureSelectXT Target Enrichment System (G7530-90000) was used for small fragment library preparation, including the following: end repair, A-tailing, adapter ligation, purification, and amplification. After construction, the quality of the sequencing libraries was assessed. Initial quantitative analyses were performed using Qubit 3.0, and the sizes of the sequencing libraries were measured by Agilent 2100. Using Bio–Rad's CFX 96 fluorescent quantitative PCR instrument and Bio–Rad's iQ SYBR Green Kit for the Q-PCR test, the concentration of each library was accurately quantified (>10 nM) to ensure quality. Each sample was 2 × 150 bp double-end sequenced with an Illumina HiSeq 4000 (Illumina, USA), and approximately 6G of raw sequencing data was output (2×). The flowchart of the experiment was shown in Figure 1.

2.4. Bioinformation analysis

The WIthin SamplE COpy Number aberration DetectOR (WISE-CONDOR) method was applied to the sequencing data of each sample.

a) First, sequence alignment was performed, and all the autosomal sequences were segmented into 'windows' with an exact length of 100 kb. Second, the GC content of each window was calculated, and the sequencing depth was subjected to Lowess correction.

b) Result judgment: The z-test was performed to obtain the z score value of the tumor tissue sample and the adjacent tissue sample for each window of the chromosomes. The windows on each chromosome are integrated together and presented as a visual image, namely, the CNV
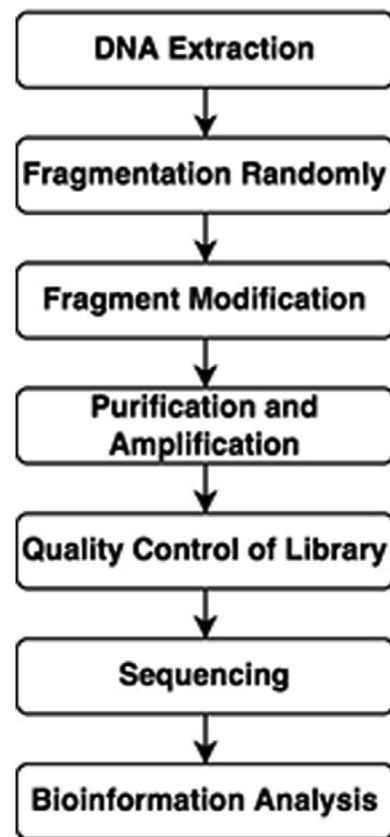

Figure 1. Flowchart of the experiment.

Autosomal atlas [11]. If the z score exceeded 3 times the standard deviation of the normal human reference data set, the window was marked as a pink square. If it exceeded 4 times the standard deviation of the reference data set, it was marked as a dark purple square, and the number of copies is represented by variations in the color shade.

c) Reference data set: One thousand normal human tissue samples collected previously were processed with the same DNA extraction, database construction, and sequencing procedures to generate the reference data set for this analysis.

2.5. Statistical analysis

The Shapiro–Wilk test was used to confirm that the data were normally distributed. For parametric data, an unpaired t test was used to evaluate differences between the two groups. For nonparametric data, differences between groups were analyzed using a Mann–Whitney U test. The chi-square test with Yates' correction and Fisher's exact test was used to compare categorical variables. A value of $P < 0.05$ was considered statistically significant. Statistical analyses were performed with SPSS software (Version 19.0, SPSS Chicago, IL, USA).

3. Results

3.1. Patient demographic features

Of the 74 samples, 28 were carcinoma tissue samples, 29 were para-carcinoma tissue samples, 13 were benign tumor tissue samples and 4 were para-benign tumor tissue samples. Para-carcinoma tissue is referred to as tumor adjacent tissue. The average age of the patients with benign nodules was 42.7 years and that of the patients with malignant nodules was 49.2 years. Of the patients with benign tissue, 35.3% (6/17) were male; 28.1% (16/57) of those with carcinoma and para-carcinoma tissue were male. Age and sex were not significantly different between the

Table 1. General and CNV information of the samples used in the study.

No.	Sex	Age	T stage*	N Stage*	Tumor size (cm)	Sample	CNV
1	Female	32	1a	0	0.7	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
2	Male	50	1b	0	1.2	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
3	Female	38	1a	0	0.8	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
4	Male	30	1a	1a	0.9	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
5	Female	48	1b	0	1.3	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
6	Female	48	1a	0	0.5	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
7	Female	51	1b	1a	1.2	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
8	Female	50	2	1a	3.2	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
9	Male	39	1b	1a	1.7	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
10	Male	40	1a	1a	0.7	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
					0.7	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
11	Female	49	1b	0	1.1	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
12	Male	39	1b	1a	1.3	carcinoma tissue	(+)
						para-carcinoma tissue	(−)
13	Female	35	1b	1a	1.2	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
14	Female	32	1a	0	0.5	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
15	Female	43	1b	1b	1.5	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
16	Female	47	1a	0	0.4	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
17	Male	33	1a	0	0.6	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
18	Female	49	1a	0	0.6	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
19	Female	47	1a	0	0.5	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
20	Female	32	1	1a	1	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
21	Male	35	1a	0	0.5	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
22	Female	63	1b	0	1.3	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
23	Female	48	1a	0	0.8	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
24	Female	34	1a	0	0.4	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
25	Female	46	1a	0	0.5	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
26	Female	67	2	0	2.7	carcinoma tissue	(+)
						para-carcinoma tissue	(+)
					6.6	benign tumor tissue	(+)
27	Female	12	1b	1a	1.9	para-carcinoma tissue	(+)
28	Female	45	1a	0	0.4	para-carcinoma tissue	(+)
29	Female	59	2	0	3.0	para-carcinoma tissue	(+)

Table 1 (continued)

No.	Sex	Age	T stage*	N Stage*	Tumor size (cm)	Sample	CNV
30	Male	43			3.7	benign tumor tissue	(−)
						para-benign tumor tissue	(−)
31	Male	52			4.9	benign tumor tissue	(+)
						para-benign tumor tissue	(+)
32	Female	31			4.7	benign tumor tissue	(−)
						para-benign tumor tissue	(−)
33	Female	36			3.4	benign tumor tissue	(−)
						para-benign tumor tissue	(−)
34	Female	68			6.5	benign tumor tissue	(−)
35	Male	52			4.7	benign tumor tissue	(−)
36	Female	45			4.2	benign tumor tissue	(−)
37	Female	52			5.2	benign tumor tissue	(−)
38	Male	59			4.0	benign tumor tissue	(+)
39	Female	55			10.4	benign tumor tissue	(−)
40	Female	42			4.2	benign tumor tissue	(+)
41	Male	55			4.2	benign tumor tissue	(+)

CNV, copy number variation.

*Based on the tumor node metastasis (TNM) staging system described in the Cancer Staging Manual of the American Joint Committee on Cancer (AJCC; 8th edition, 2018).

malignant and benign groups (p = 0.10; p = 0.29). The size of malignant nodules was significantly smaller than that of benign nodules (1.1 cm vs. 5.0 cm, P < 0.001). None of the patients showed distant metastasis, and the tumor node metastasis (TNM) staging of the patients is shown in Table 1.

3.2. CNV portfolio in PTC

In the CNV mapping of the carcinoma tissue samples, pink or purple rectangles, which mark the locations of the CNVs, could be clearly found near the end of chromosome 5q, near the centromere of chromosome 7, almost at the end of chromosome 8q, at the middle of chromosome 10p, and near the end of chromosome 17. In detail, the CNVs were located at 150–154 Mb (chr. 5), 68–74 Mb (chr. 7), 133–139 Mb (chr. 8), 20–24 Mb (chr. 10) and 58–65 Mb (chr. 17). However, the CNV portfolio could not be identified in the mapping of the most benign tumor tissues. Figures 2 and 3 show the classic autosomal CNV analysis mapping of PTC samples (positive) and benign thyroid tumor samples (negative).
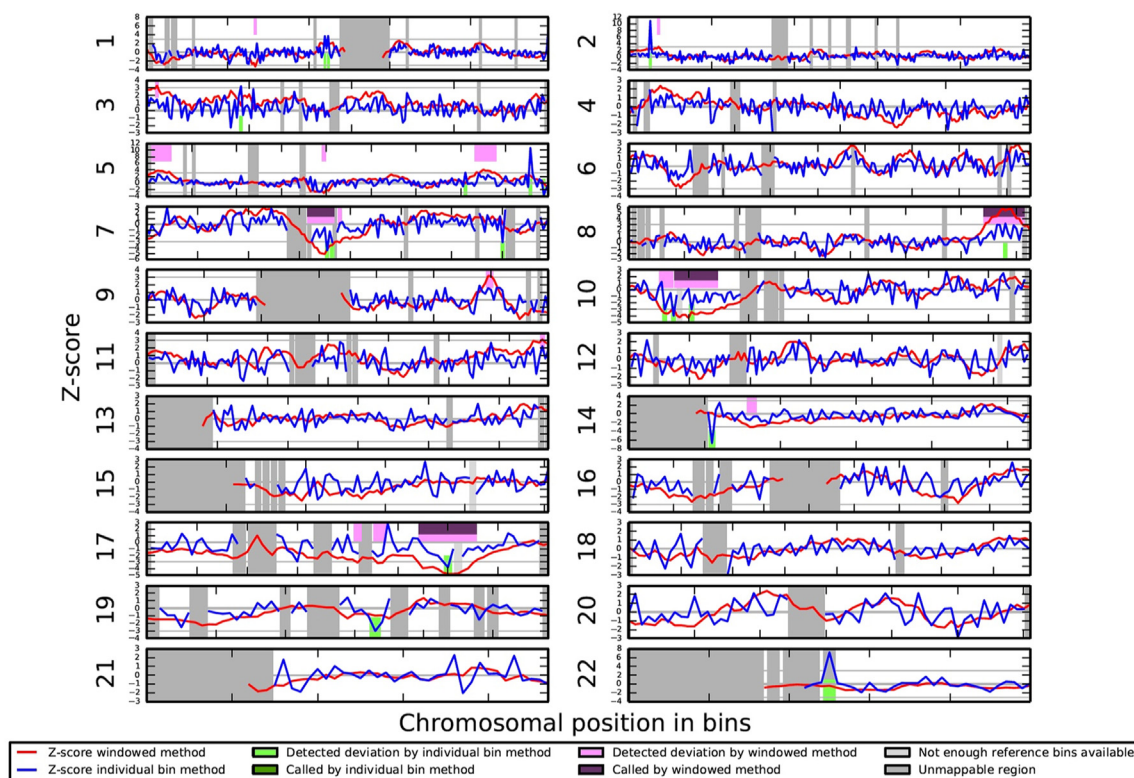

Figure 2. The autosomal CNV analysis mapping of patient No. 4 in Table 1. A 30-year-old man with a 0.9 cm PTC. Pink and purple rectangles (in red square frames) are shown at particular locations on the chromosomes. The map showed that chr. 8 demonstrated an abnormal increase in copy number, and chr. 7, chr. 10, and chr. 17 demonstrated an abnormal decrease in copy number.
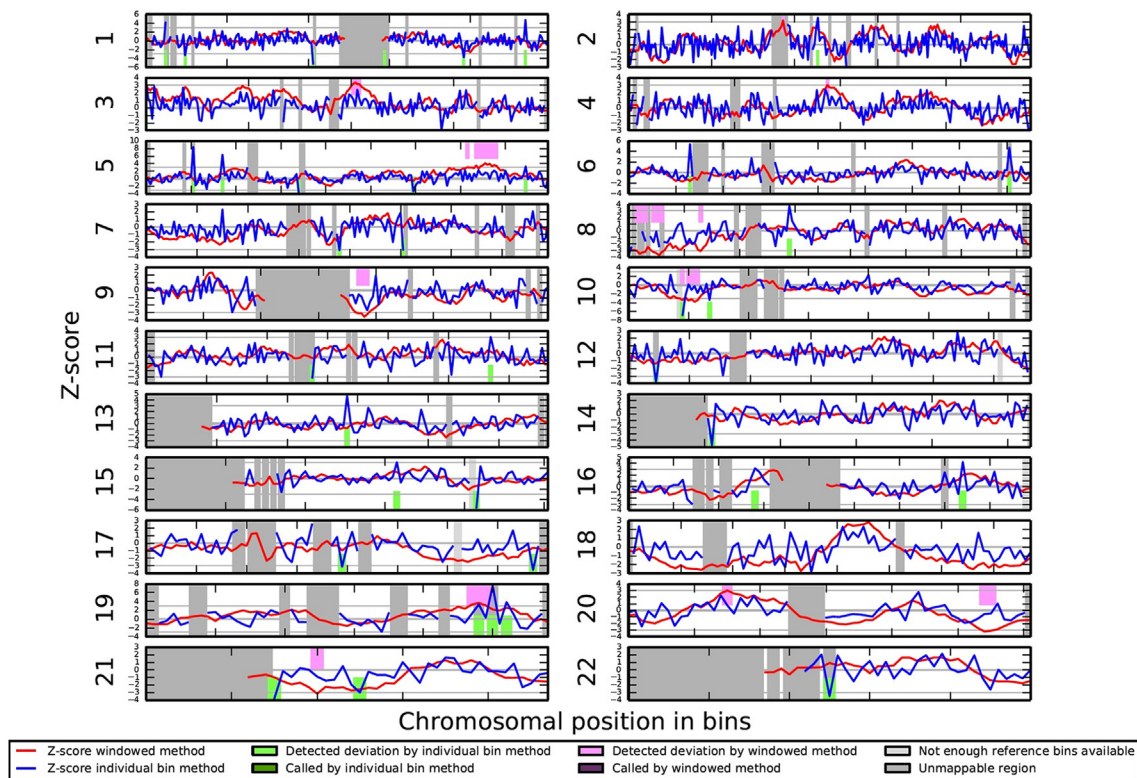
Figure 3. The autosomal CNV analysis mapping of patient No. 32 in Table 1. A 31-year-old woman with a 4.7 cm nodular goiter. No CNV regions (in red square frames) were identified.
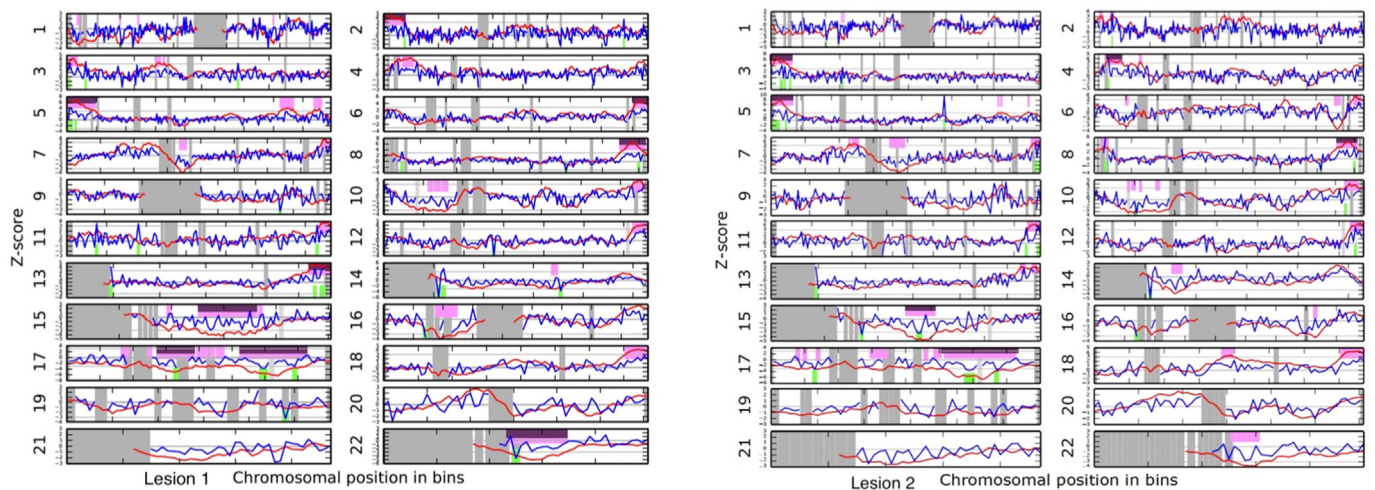

Figure 4. The CNV portfolios of two different lesions from the same person (patient No. 10, a 40-year-old man with PTC).

3.3. Identification rules of positive and negative samples

We categorized the five CNV regions into two groups: group A (including chr. 5 and chr. 8, abnormal increase in copy number) and group B (including chr. 7, chr. 10, and chr. 17, an abnormal decrease in copy number). If at least one CNV in group A and at least two CNVs in group B were identified, the sample was considered 'positive', indicating that the sample was malignant. Otherwise, the sample was considered 'negative'.

3.4. The diagnostic value of the CNV portfolio

According to the above rules, the test results of all samples were obtained and are summarized in Table 1. All 28 PTC samples were

positive, and the CNV patterns were very similar. Moreover, the para-carcinoma tissue samples from the same patient also showed positive results except for in one case (3.6%), which was negative. For the 12 patients with benign thyroid tumors only, a total of 16 samples, including benign tumor tissues and adjacent tissues, were collected; 11 of them tested negative (68.7%), and the others tested positive (31.3%). The only benign nodule sample from a patient with PTC also tested positive. The CNV portfolios of different lesions from the same person were almost the same [Figure 4], illustrating that the CNV test method had good reproducibility.

For tumor tissue, the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and area under the curve (AUC) of the test method were 100%, 66.7%, 87.5%, 100.0%,

90.0% and 0.83 (95% confidence interval [CI], 0.67–1.00), respectively. For para-tumor tissue, the sensitivity, specificity, PPV, NPV, accuracy and AUC were 96.6%, 75.0%, 96.6%, 75.0%, 93.9% and 0.86 (95% CI, 0.60–1.00), respectively.

The results also showed that the partial deletion of chromosome 22 was common (14/28) in cancer samples, and there was a certain probability of complete absence (8/28), amounting to 78.6%.

4. Discussion

According to our findings, thyroid nodules can be diagnosed by the presence of CNVs at the corresponding positions on chromosomes 5, 7, 8, 10, and 17. All malignant tumors can be diagnosed by the CNV map with a sensitivity of 100%, suggesting that using the novel molecular marker to screen thyroid puncture samples or surgical samples will not lead to missed diagnoses. Moreover, the similarity of the CNV patterns of all 28 PTC samples indicated that PTC may have arisen from the molecular mechanism in all of the patients. It should be noted that due to the low-depth sequencing strategy, these five CNVs may not all be present in a single test because the number of reads within a certain region may be insufficient to show significant statistical deviations.

Molecular variations, which often occur before morphological changes in abnormal cells, can be detected by microscopy. In one case, 0.1-cm cancerous foci that were barely captured either macroscopically or histologically tested significantly 'positive' at the molecular level. Moreover, the benign nodule sample of this patient also showed 'weakly positive' results. This finding suggests that cancerous signals could be greatly amplified even when morphological aberrations appear indistinct microscopically. This can also explain why most para-carcinoma tissues from patients with PTC tested positive, and their cancerous signals were almost the same but slightly weaker than those of carcinoma tissues from the same person.

Some patients with benign tumors had a positive test result. Although we know that molecular tests are more sensitive than pathologic markers, it is difficult to determine whether the specificity of this test is insufficient or whether the pathological marker lags behind. For these patients, further tracking studies are required. It was noted that only 3.6% of the paracancerous tissues tested negative, illustrating that the method does not require strict sampling accuracy. The puncture procedure may correct sampling mistakes to some extent.

This mechanism associating these CNVs and thyroid cancer was partially revealed by genome database GenBank analysis (https://www.ncbi.nlm.nih.gov). The abnormal increase in chromosome 8 occurs in the 133–139 Mb segment containing the thyroglobulin (TG) gene. TG is a tumor marker of differentiated thyroid cancer (DTC). TG gene copy number gain leads to the overexpression of downstream proteins and increases the risk for locoregional recurrence and distant metastases in thyroid cancer patients [12]. This segment also includes the NDRG1 gene, encoding N-myc downstream regulatory factors that lead to the proliferation and apoptosis of tumor cells. Previous studies have shown that upregulating the expression of the NDRG1 gene leads to poor prognosis and promotes the progression of bladder cancer [13]. The reduced copy number of chromosomes 17 occurs in the 58–65 Mb segment that contains the BRIP1 gene, which encodes BRCA1-interacting protein c-terminal helicase 1 and is involved in chromosomal damage repair. Several studies have suggested that BRIP1 gene mutation leads to an increased risk for developing multiple cancers [14, 15]. Chromosomal damage repair cannot be executed properly due to gene deletion or site-specific mutations, which may result in the downregulation of proteins or loss of biological function.

Previous studies have found deletions on chromosome 22 in thyroid follicular adenoma and follicular variants of PTC, which were used for screening thyroid cancer [16, 17]. In this study, although partial deletion of chromosome 22 in cancer samples was common (14/28) and there was a certain probability of complete absence (8/28), it was

not a suitable marker for thyroid cancer screening because of its sensitivity.

Finally, we also found different molecular characteristics among the pathological subtypes. All four follicular variants of PTC (not specified in Table 1) had abnormal copies in the segment of chromosome 15, which indicates that the CNV may play an important role in the evolution of cancer. The CNV portfolio is likely to be a promising molecular tool for the categorization of PTC.

5. Conclusion

Thyroid nodules can be diagnosed by screening for CNVs at the corresponding positions on chromosomes 5, 7, 8, 10, and 17. High-volume CNV sequencing analysis can greatly increase the identification rate of PTC, helping to avoid delays in further treatment.

Declarations

Author contribution statement

Xingjian Lai, Xiequn Xu: Conceived and designed the experiments; Analyzed and interpreted the data.

Luying Gao: Analyzed and interpreted the data; Wrote the paper.

Gaoying Zhou, Jinhui Wang: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data.

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] L. Hegedüs, Clinical practice. The thyroid nodule, N. Engl. J. Med. 351 (2004) 1764–1771.

[2] B.R. Haugen, E.K. Alexander, K.C. Bible, et al., American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer, Thyroid 26 (2015) 1–133.

[3] P. Valderrabano, B. McIver, Evaluation and management of indeterminate thyroid nodules: the revolution of risk stratification beyond cytological diagnosis, Cancer Control 24 (2017)1073274817729231.

[4] M. Decaussin-Petrucci, F. Descotes, L. Depaepe, V. Lapras, et al., Molecular testing of BRAF, RAS and TERT on thyroid FNAs with indeterminate cytology improves diagnostic accuracy, Cytopathology 28 (2017) 482–487.

[5] A.C. Insilla, A. Proietti, N. Borrelli, et al., TERT promoter mutations and their correlation with BRAF and RAS mutations in a consecutive cohort of 145 thyroid cancer cases, Oncol. Lett. 15 (2018) 2763–2770.

[6] D. Pinkel, D.G. Albertson, Array comparative genomic hybridization and its applications in cancer, Nat. Genet. 37 (Suppl) (2005). S11–7.

[7] F. Kaveh, L.O. Baumbusch, D. Nebdal, et al., A systematic comparison of copy number alterations in four types of female cancer, BMC Cancer 16 (2016) 913.

[8] S.E. Schumacher, B.Y. Shim, G. Corso, et al., Somatic copy number alterations in gastric adenocarcinomas among Asian and Western patients, PLoS One 12 (2017) e0176045.

[9] H.G. Woo, J.H. Choi, S. Yoon, et al., Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer, Nat. Commun. 8 (2017) 839.

[10] D.J. VanderWeele, R. Finney, K. Katayama, et al., Genomic heterogeneity within individual prostate cancer foci impacts predictive biomarkers of targeted therapy, Eur. Urol. Focus 5 (2019) 416–424.

[11] R. Straver, E.A. Sistermans, H. Holstege, et al., WISECONDOR: detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme, Nucleic Acids Res. 42 (2014) e31.

[12] J.D. Lin, Thyroglobulin and human thyroid cancer, Clin. Chim. Acta 388 (2008) 15–21.

[13] A. Li, X. Zhu, C. Wang, et al., Upregulation of NDRG1 predicts poor outcome and facilitates disease progression by influencing the EMT process in bladder cancer, Sci. Rep. 9 (2019) 5166.

[14] S.H. Jung, M.S. Kim, C.K. Jung, et al., Mutational burdens and evolutionary ages of thyroid follicular adenoma are comparable to those of follicular carcinoma, Oncotarget 7 (2016) 69638–69648.

[15] W. Jian, K. Shao, Q. Qin, et al., Clinical and genetic characterization of hereditary breast cancer in a Chinese population, Hered. Cancer Clin. Pract. 15 (2017) 19.

[16] Y. Liu, L. Cope, W. Sun, et al., DNA copy number variations characterize benign and malignant thyroid tumors, J. Clin. Endocrinol. Metab. 98 (2013) E558–E566.

[17] L. Ye, X. Zhou, F. Huang, et al., The genetic landscape of benign thyroid nodules revealed by whole exome and transcriptome sequencing, Nat. Commun. 8 (2017)15533.