

Gut microbiome is associated with thyroid nodule and functions

Suying Ding (✉ fccdingsy@zzu.edu.cn)

First Affiliated Hospital of Zhengzhou Univeristy <https://orcid.org/0000-0002-2649-8332>

Ang Li

Zhengzhou University First Affiliated Hospital

Tiantian Li

Zhengzhou University First Affiliated Hospital

Xinxin Gao

Zhengzhou University First Affiliated Hospital

Hang Yan

Zhengzhou University First Affiliated Hospital

Yu Wang

Zhengzhou University First Affiliated Hospital

Jingfeng Chen

Zhengzhou University First Affiliated Hospital

Mengfan Jiao

Zhengzhou University First Affiliated Hospital

Zhigang Ren

Zhengzhou University First Affiliated Hospital

Qingmiao Shi

Zhengzhou University First Affiliated Hospital

Meng Huang

Zhengzhou University First Affiliated Hospital

Lin Wang

Zhengzhou University First Affiliated Hospital

Jingmei Li

Zhengzhou University First Affiliated Hospital

Yihan Dong

Zhengzhou University First Affiliated Hospital

Zhuqing Jia

Zhengzhou University First Affiliated Hospital

Lihua Zhang

Zhengzhou University First Affiliated Hospital

Qing Feng

Zhengzhou University First Affiliated Hospital

Huandi Han

Zhengzhou University First Affiliated Hospital

Xiang Yang

Zhengzhou University First Affiliated Hospital

Yang Li

Zhengzhou University First Affiliated Hospital

Benchen Rao

Zhengzhou University First Affiliated Hospital

Detao Yin

Zhengzhou University First Affiliated Hospital

Hongqiang Li

Zhengzhou University First Affiliated Hospital

Runsheng Ma

Zhengzhou University First Affiliated Hospital

Qiang Zeng

Chinese PLA General Hospital

Zujiang Yu

Zhengzhou University First Affiliated Hospital

Research

Keywords: Gut microbiota, Thyroid nodule, Thyroid function, Metagenomics, Nutrition metabolism

Posted Date: March 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16228/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Thyroid nodule affects nearly half of the adult population. Accumulating evidence suggests that gut microbiota plays an important role in the thyroid disorders and functions, yet the association between gut microbiota capacity and thyroid nodule and function have not been studied comprehensively.

Methods: We performed a gut microbiota whole-genome wide association study in 537 thyroid nodule participants and 725 controls from the general Chinese population. Thyroid nodules were with the Thyroid Imaging Reporting and Data System assessment.

Results: Participants with the highest grade thyroid nodule had decreased gut microbial species and functional richness compared with controls and thyroid nodule with lower grade. Microbiota composition of the low-grade thyroid nodule was similar to that in controls rather than high-grade thyroid nodule. We found significant alternations of overall microbial functional composition in high-grade thyroid nodule but not the species composition, only a small portion of microbial species were associated with high-grade thyroid nodule. The high-grade thyroid nodule gut microbiome was characterized by greater amino acid degradation and glycolysis, also with fewer *Butyrivibrio* and lower butyric acid production. Thyroid functions were positively associated with L-histidine and L-tyrosine related biosynthesis pathways, which are essential components of thyrotropin-releasing hormone and thyroxine, also correlated with broad nutrition metabolism pathways including fatty acids, B family vitamins and nucleotides. High-grade thyroid nodule discrimination models based on the gut microbiome and demography were validated using 294 independent samples and showed stable classification power.

Conclusion: Our study comprehensively described the gut microbiome characteristics in thyroid nodule and suggested potential role of the key gut microbiota-driven functional pathways in thyroid nodule and functions metabolism.

Background

Thyroid nodule (TN) is in increasing incidence worldwide, such that they are now present in 20–76% of the adult population.^[1] Although the majority of TNs are asymptomatic and benign, 15% were found to exhibit significant increases in size during a 5-year follow-up^[2] and 5–15% can be malignant^[3], making this the type of cancer that is most rapidly increasing in incidence globally^[4]. Age, sex, the metabolic syndrome, and iodine intake are risk factors for the development of TNs^[5, 6], and several human genome loci, including TRPM3, EPB41L3, and AP005059 have been found to be associated with TNs^[7]. The importance of thyroid function for appropriate metabolic regulation has been recognized for decades, but this is affected by numerous environmental factors^[8] and is partially genetically determined in humans^[9]. However, the aetiology of TNs and the role of genetics in thyroid function remain to be fully elucidated.

Evidence is accumulating that the gut microbiome plays an important role in thyroid disorders and function by regulating mineral metabolism and enterohepatic cycling^[10]. The importance of gut microbiota on human health has been underlined for more than one decade, however, our understanding of its role in common conditions such as thyroid nodules and thyroid functions are still very limited. A recent study showed gut microbial dysbiosis in 18 patients with TN and associations between the microbiome and thyroid function^[11]. However, the significance of these findings was restricted by the small number of patients studied and a suspicion of batch bias^[12], which might have been caused by the use of 16S ribosomal RNA (rRNA) gene amplicon sequencing.

For the present study, we recruited 537 participants with TN and 725 controls as a primary analysis cohort from the Chinese Healthy Integrative-omics project (CHIP). 670 of these participants underwent serum thyroid function testing, including the measurement of free triiodothyronine (FT3), free thyroxine (FT4), and thyroid stimulating hormone (TSH) concentrations. In addition, whole genome shotgun sequencing was performed on stool samples. Using these data, we aimed to define the gut microbial characteristics associated with TNs and thyroid function.

Results

Characteristics of the primary cohort

The primary cohort included 1,262 participants (38.43% women) after all exclusions (see method for details). In 537 (42.55%) participants at least one thyroid nodule was detected. The mean ages of the control and TN groups were 42 and 47 years old. The participants were then classified into four groups according to the thyroid ultrasonographic findings: (1) 725 demonstrated no remarkable findings (controls); (2) 172 had Thyroid Imaging Reporting and Data System (TI-RADS) level 2 (TN2); (3) 328 had TI-RADS level 3 (TN3); and (4) 34 had TI-RADS level 4a and 3 had TI-RADS level 4b (TN4), of which six TI-RADS level 4a and 3 TI-RADS level 4b underwent fine-needle aspiration biopsy, which did not identify malignancy. TI-RADS 4b were not included in the downstream analysis, despite having no malignancy. Participants with more than one thyroid nodule were classified according to the highest TI-RADS level. The demography and laboratory test data are shown in **Table S1**.

Age and blood pressure were higher in the TN group. In addition, when compared with female controls, female TN had higher body mass index (BMI), waist and hip circumference (WC and HC), triglyceride (TC), total cholesterol (TG), high-density lipoprotein-cholesterol (HDL-C) concentrations, and red blood cell counts (RBC). No significant differences were found between the TN2 and control groups, or between the TN3 and TN4 groups, except for higher systolic blood pressure in female TN4 and higher RBC in male TN4. Thyroid function and lifestyle habits, including smoking and drinking, were not related to TN incidence. Principal component analysis (PCA) based on demographics identified the principal components PC1 and PC2, which correlated with TI-RADS. Pulse pressure (PP) and age were with accordant contribution, together with BMI, WC and HC (**Figure S1**).

Gut microbial characteristics of the primary cohort

A total of 928 species presented in at least five samples were used for downstream analysis. 343 species of these were viruses, six were Archaea, and 575 were bacteria. 61 species had a median relative abundance >0.01%, comprising 76.4% the total microbiome. 839,310 gene families and 459 metabolic pathways were represented in at least five samples (**Figure S2**). The representation of metabolic pathways was more equal than that of species, demonstrated by mean Gini indexes of 0.768 and 0.983, respectively (**Table S2**). In addition, the intra-group and inter-group similarities and distances standard deviations were lower, with respect to the metabolic pathways represented, compared with the similarities based on species composition. (**Figure S3**). Principal coordinates analysis (PCOA) conducted on the basis of metabolic pathway composition showed that the principal coordinates PCO1 and PCO2 correlated with the presence of TNs and TI-RADS (**Figure 1a**), but not species composition.

TN2 were more similar to controls when inter-group distances were compared with TN3 and TN4 (**Figure S4**), except with respect to Spearman coefficient distance (SCD) based on species composition. When the distribution of PCOs was analysed, few significant differences were found between the control and TN2 groups (**Table S3 and Table S4**). There were significant differences between TN2 and TN3 in PCO1 for metabolic pathways (**Table S4**). Because >99% of TNs with TI-RADS level 2 are benign and do not require treatment^[13], we then classified control and TN2 as having no or low grade TIRADS (Group LTN) and those with TN3 or TN4 as having advanced grade TIRADS (Group HTN).

1259 individuals could be clustered into two enterotypes on the basis of species composition (**Figure S5**). Enterotype 1 had higher microbial diversity, and contained a higher proportion of women (41.1%, compared with 30.9% in enterotype 2). None of TN incidence, TI-RADS level, or thyroid hormone concentrations (serum FT3 and FT4) were significantly associated with enterotype, except that men with enterotype 1 had higher serum TSH concentrations (**Figure S6**).

Lower microbial richness in TI-RADS level 4a

The number of observed species (OBS) was significantly lower in TN4 than in the other groups (Wilcoxon's rank sum test, $p = 0.005$, $p = 0.01$, and $p = 0.011$, compared with control, TN2, and TN3, respectively). 56.91% (247 of 434) of the species not detected in TN4 were viruses, and these were less abundant than other species in all the participants: the median fold-differences (common species *versus* species not found in TN4), based on the relative abundance in all the participants (including TN4), were 121.52 and 2.67 for bacteria and viruses, respectively. The majority of the viruses not detected in TN4 were phages that might be hosted by *Enterobacteriaceae*. The number of gene families identified and viruses ratio were also lower in TN4 (**Figure 1b-e**).

The OBS was significantly higher in women in the control and TN2 groups, and was also associated with the age of the women and the BMI of the men in the control group (**Figure S7**). In considering female ratio, BMI and age were higher in the TN4 group than in the control and TN2 groups. In addition, the TN4 group was smaller; therefore, we used bootstrap sampling to generate sample number-, sex-, age- and

BMI-matched subsets of controls and TN2, such that each subset contained 40 randomly chosen samples. However, the OBS in the TN4 group was still significantly lower after resampling (Wilcoxon's rank sum test, median $p = 0.0365$ and median $p = 0.0165$, compared with controls and TN2, respectively) (**Figure S8**). We also used analysis of variance (ANOVA) to confirm the lower microbial richness in the TN4 group, after adjusting for age, sex, and BMI ($p = 0.005$, $p = 0.014$, and $p = 0.017$, compared with the control, TN2, and TN3 groups; **Table S5**). No ultrasonographic parameters were associated with OBS, except that male TN without thyroid echo had higher OBS ($p = 0.004$) (**Figure S9**).

Gut microbiota changes in thyroid nodules

We analysed 418 prokaryote species and 440 metabolic pathways that had been identified in ≥ 20 samples. Sixteen species and 24 pathways withstood adjustment for multiple covariates (sex, age, and BMI) and false discovery rate (FDR) controlling multiple testing to show distinct relative abundances between LTN and HTN, by using multivariate association with linear models (MaAsLin2, see method, q value < 0.25) (**Figure 2a and Table S6**), and these comprised 0.67% and 4.68% of the total numbers of microbial species and pathways, respectively (**Figure S10**).

We found that butanoate formation pathways (*P163-PWY* and *PWY-5676*) and *Butyrivibrio* (unclassified) were over-represented in LTNs (**Figure 2c**), as was the essential transfer RNA charging pathway (*TRNA-CHARGING-PWY*), which incorporates amino acids into growing polypeptides, could promote butyrate formation by the intestinal microbiota^[14,15]. L-isoleucine biosynthesis (*PWY-3001*) was enriched in HTNs, and *in vitro* studies of ruminal fermentation have shown that butyrate production increases with isoleucine addition^[16].

The metabolic pathways that were perturbed in HTNs were mainly involved in amino acid degradation and glycolysis. Several metabolomic studies have shown the accumulation of L-glutamate^[17] and lactate^[18,19] in benign cases of TN, which might be contributed to by the L-histidine and 4-amino butanoate (GABA) degradation (by *PWY-5030*, *HISDEG-PWY* and *GLUDEG-I-PWY*), glycolysis (*Glycolysis-I*, *PWY-5484*) and homolactic fermentation (*ANAEROFrucat-PWY*) pathways (**Figure 2b**). The human intestinal microbiome can ferment lactate to generate butyrate in the presence of polysaccharide^[20]. However, the accumulation of lactate and the resulting low pH environment would inhibit butyrate formation^[21,22]. In addition, inter-individual variations in short-chain fatty acid (SCFA) synthesis ability are typical^[23]. Although L-glutamate degradation (*P162-PWY*) could also yield butyrate *via* butanoyl-CoA, *P162-PWY* was far less abundant (mean relative abundance 0.0077%) than *PWY-5676* (0.108%, detected in 867 LTN and 341 HTN), as was another LTN-enriched butyrate synthesis pathway (*P163-PWY*; 0.00018%) and a neutral pathway *CENTFERM-PWY* (0.012%) that can also ferment pyruvate to produce butanoate (**Figure S11**). The gut microbiota participating in *PWY-5676*, which was the main butanoate produce pathway via acetyl-CoA fermentation, could not be classified into any known species, but were positively associated with several important butyrate-producers (**Table S7 and Figure S12**). As a well-recognized histone deacetylase inhibitor^[24,25], sodium butyrate could induce growth arrest in a variety of

normal cell types^[26,27], enhance radiosensitivity^[28], and regulate the expression of several tumour-suppressor genes^[29,30,31], as well as increasing I¹²⁵ uptake^[32], which has been shown in thyroid follicular carcinoma cells to occur by restoring human sodium-iodide symporter function (nHIS). It has also been shown to increase thyroid hormone receptor expression in GH3^[33] and glial C6^[34] cells.

Furthermore, we identified *Bacteroides plebeius*, which is of marine origin and was found to be ubiquitous in a Japanese cohort^[30] and with porphyrin-degrading ability^[31], to be enriched in male LTNs after covariates (BMI and age) adjusting (multiple linear regression, adjusted $p < 0.05$). *B. plebeius* positively correlated with BMI^[35,36], despite LTNs having lower BMI. Numerous *in vitro* studies have demonstrated that red or brown seaweeds can be fermented to produce butyrate in the human gut, but porphyrin or green seaweeds are not effective as substrates^[37]. Increases in the populations of *Bifidobacterium* and *Lactobacillus* were accompanied by the loss of butyrate-producing bacteria in patients with inflammatory bowel disease^[38]. We found that the populations of two *Lactobacillus* species (*L. salivarius* and *L. sanfranciscensis*) were higher in whole HTNs (see method, $q < 0.25$) and that *Bifidobacterium longum* was enriched in female TN4 (multiple linear regression, adjusted $p < 0.05$, **Table S8 and Figure S13**).

Association of gut microbiota and thyroid function

Data from 624 participants with normal thyroid function were used for association analysis. All the thyroid parameters measured correlated with age and sex (**Table S9**). 200 pathways and ten species were associated with at least one thyroid parameter after adjustment for covariates by using multivariate association with linear models (MaAsLin2, see method, q value < 0.25) (**Table S10**). TSH inversely correlated with FT3 and FT4 concentrations, and the microbial features that positively correlated with TSH also showed inverse associations with FT4 and FT3, but for the majority no significant associations were identified (**Figure 3a**).

Thyroid function was associated with a broad range of metabolic pathways (**Figure 3b**). L-histidine biosynthesis (HISTSYN-PWY) was positively associated with serum TSH, perhaps because it is an essential component of thyrotropin releasing hormone (TRH) and essential amino acid that is not synthesized [de novo](#) in humans, and therefore could increase TSH synthesis and secretion from the anterior pituitary^[39]. Tyrosine biosynthesis (*PWY-6629*, *PWY-6630*) and biosynthesis (*PWY-6628*) were also positively associated with FT3, tyrosine is composed of iodinated tyrosine residues (**Figure 3c**), phenylalanine is an essential amino acid which could be converted into tyrosine via [hydroxylation](#) in human. Previous studies have shown that low serum tyrosine is associated with reduced circulating FT3 concentration and thyroid dysfunction^[40,41,42]. Otherwise, the majority of amino acid and adenosine biosynthetic pathways negatively correlated with FT3. Caloric restriction could reduce thyroid hormone concentrations by inhibiting 5'-deiodinase activity, which might be caused by lower production of NADPH (*PWY-7269*)^[9], accompanied by lower ATP content and synthase activity^[43,44]. Conversely, administration of thyroid hormones could reduce plasma total free amino acid concentration^[45] and inhibit ATP hydrolysis^[46].

Ubiquinol (Coenzyme Q, CoQ), menaquinone (Vitamin K2), various B vitamins (including B9, B12, B1, and B7) which are only synthesized by plants, yeasts and bacteria but not by human host, and some essential pathways involved in CoQ and vitamin biosynthesis were positively associated with FT3 (**Table S10**). We found Biotin (vitamin B7), which is a commonly used supplement in practice, might induce abnormal thyroid function test results^[47], *via* its direct effect to increase thyroid hormone and reduce TSH concentrations ^[48,49].

Models to distinguish advanced TIRADS thyroid nodules

To evaluate the potential for the gut microbiome to be used to distinguish HTN, we constructed two types of classification models: a) Demographic models based on age, gender, BMI and SP. b) Mixed models based on the microbial species present and their metabolic pathways, together with the demographics for 1,259 samples. All the available 180 LTNs and 114 HTNs from the CHIP cohort were used as an independent validation cohort; the corresponding samples were still in sequencing while we were analyzing the primary cohort data. No obvious batch effect bias was detected between the train and the independent validation cohorts, so we did not filter out any samples (**Figure 4a**). Model performance was assessed by receiver operating characteristic (ROC) curve analysis using the area under the curve (AUC).

We used a combination of gradient boosting machine (GBM) feature importance ranking and random forest backward elimination to remove irrelevant features that might have hampered model performance in a 10-fold cross-validation in train samples (**Figure S14**), then randomly divided 1,259 samples into 886 train and 373 cross-validated samples. Models constructed on train samples performed the best with 53 selected features (**Figure S15 and Table S11**), which achieved an AUC of 0.734 (95% confidence interval 0.678–0.789) and AUC of 0.739 (0.672-0.806) on cross validate samples and independent validate samples (**Figure 4c-f**). Demographic models did not filter out any features and could achieve AUC of 0.7016 (0.645-0.7581, $p = 0.086$) and 0.6867(0.6175-0.7559, $p = 0.039$) compared with mix models. Age, gender, BMI and SP still had significant contribution to mix models and 10 selected microbiota features were significantly associated with HTN group (**Figure 4b**). A generalized AUC of 0.736 was obtained after iteratively choosing 200 random samples from the independent cohort 100 times (**Figure 4g**).

Discussion

TN is the most common watchful waiting endocrine disorder in clinical practice, normally it is not considered as “sick” but with complex and unknown aetiology. Here, we described the gut microbial characteristics of patients with TN and their relationships with thyroid function using a large number of samples collected from a Chinese cohort. Low-grade TN (TN2) were similar in their demographics and gut microbiota characteristics to controls, and they were therefore analysed together as the LTN group, which is appropriate considering that nearly 100% of TN2 nodules are benign and require no treatment. We found that the presence of highest grade TN (TN4), which was with undetermined malignancy risk (5–10%)^[50] was associated with lower gut microbiota species and function richness, though most of vanished species in TN4 might be passenger microbiota with low relative abundance. Overall microbiota

function composition of HTN was altered but not at the species level, only a small proportion of the gut microbiota species were associated. Zhang^[11] and colleagues reported increased microbiota species richness in TN and dramatic alternations in the overall microbial structure based on 16 s rRNA gene sequencing and operational taxonomic units (OTU) profiling. TI-RADS or other thyroid nodule assessments were not available and it is hard to compare these inconstant findings based on different sequencing methods.

We identified the accumulation of key metabolites including glutamate and lactate by gut microbes in HTNs, which were constant with previous metabolomics studies^[17, 18]. Our findings indicated that lower butyrate production by gut microbes might be associated with HTN because of the resulting capability to increase iodine-125 uptake in thyroid follicular carcinoma cells^[32], iodine deficiency is one cause of thyroid nodules development^[6], however no similar evidence were showed in normal cell lines. In addition, we did not identify specific butyrate formation microbiota reduced in HTN but an unclassified *Butyrivibrio* specy, and it was not significantly associated with detected butyrate production pathways, thus reduced butyrate formation capacity might be from non-specific population. *Bifidobacterium longum* had been recognized as a probiotic in the general population, was more common in female TN4, and therefore should be administered more cautiously, because it can reduce the population of butyrate-producing microbes³⁸.

Various nutrient biosynthesis pathways were associated with thyroid function, indicating the importance of gut microbiota for metabolism^[51, 52]. L-histidine and L-tyrosine biosynthesis pathways were positively associated with serum thyroid hormone concentrations, and are essential components of TRH and thyroxine. In addition, high-dose biotin (vitamin B7) were used to increase thyroid function in the clinic^[48]. Hypothyroidism is associated with vitamin D^[53], B6, and B12^[54] deficiency and vitamin supplementation might ameliorate hypothyroidism by promoting levothyroxine uptake^[55], also gut commensals are highlighted as roles of reservoir producers for B family vitamins which could not be synthesized by human host^[56, 57], but no evidence showed these vitamin supplementary improved thyroid function in normal people. Finally, discriminative diagnostic models based the composition of the gut microbiome and demography showed stable discriminative power in an independent validation cohort, also were models based only on demographic features, although ultrasonic elastography should remain the first choice method for TN diagnostic screening^[50].

Limitations of this study should be mentioned. Those associations identified are not specific and causal, for example, butyrate metabolism has also been shown to be associated with several disorders including type 2 diabetes, inflammatory bowel disease, and obese^[58]. In addition, findings will still at least partially be confounded by demographic parameters even after strict adjustment.

Conclusion

In conclusion, the present study described gut microbiota characteristics in thyroid nodule and its relationship with human serum thyroid functions. Our findings indicated alternations in gut microbiota functions in high-grade TN but not species composition. Notably, we found several gut microbiota driven metabolism pathways were associated with thyroid nodule and functions, which directly participate in metabolites biosynthesis, importantly, these metabolites are with capability of cell disturbance might be associated with thyroid nodule and are essential components of thyroid hormones, and the majority could not be synthesized de novo in humans. Although these associations were not causal and needed further evaluation, we have demonstrated that the gut microbiome is linked to high grade TN and thyroid functions, relationships that are likely to be mediated via microbial nutrition metabolism, which suggested its broad impact on human common conditions.

Materials And Methods

Study population

The participants in this study were recruited from the Chinese Healthy Integrative-omics project (CHIP), which was approved by the Ethics Committee from the First Affiliated Hospital of Zhengzhou University. Eligibility criteria required individuals who were all over 18 years old to have undergone routine physical examinations and volunteered to join CHIP cohort with informed content, at the Physician Health Center of the First Affiliated Hospital of Zhengzhou University, Eastern District of the Hospital, from January 2018 to now. Pregnant women and individuals who were under 18 were excluded. The participants abstained from alcohol three days before the physical examination, fast at 8:00 pm and were water-free after midnight the previous evening. Participants were allowed to take medicine with water not exceeding 200ml before drawing blood in the morning. Plavix and anticoagulant drugs including warfarin and aspirin were stopped a week before. For 1389 participants underwent thyroid ultrasound scan, 69 with antibiotics usage in the last three months were excluded, 53 participants with thyroid surgery, thyroxine agent intake (all are levothyroxine sodium), thyroid enlargement or lymphadenectasis were excluded. In addition, 5 with TN because of lacking TI-RADS assessment were excluded.

Sample collection

Stool samples were obtained from recruited participants. Samples were immediately stored at -20°C and frozen at -80°C within 30min before transport to the laboratory. Blood and stool samples were all obtained on the same day at the hospital before 10 am. All blood samples were drawn in the morning before 10:00. During blood drawing and fecal retention, participants maintained fasting.

Laboratory test

Lavender-Top tubes that contained EDTA as an anticoagulant were used to test the blood routine examination with 1-2ml of venous blood and the tubes were shaken up and down evenly 5-8 times to prevent blood from clotting. Yellow-Top tubes which contained separating gel and coagulant to accelerate coagulation were used for blood lipids blood glucose and liver function with 2ml venous

blood. Blood samples were detected using a Roche cobas 8000 automatic biochemical analyzer. Red-Top tubes without anticoagulant were used for thyroid function with 2ml venous blood, including free triiodothyronine (FT3), free thyroxine (FT4) and thyroid stimulating hormone (TSH). After drawing blood, all the tubes were placed in specimen boxes. The collected blood samples were entered into the hospital specimen transmission system and sent to the clinical central laboratory by a logistics trolley within 20 minutes. The normal range for FT3, FT4, TSH are 3.28-6.47 picomoles per liter of blood (pmol/L), 7.9-18.4 picomoles per liter of blood (pmol/L), 0.34-5.6 milliunits per liter of blood (mU/L), respectively. The standard of the normal range was our hospital's built-in cutoff. UniCel Dxl 800 Immunoassay System from Beckman Coulter was used to test the blood samples.

Demographic Feature Measurement

Participants stood on the footprint of the scale with bare feet and face the scale to measure height and weight. After hearing the voice prompts, participants kept their eyes level, chest straight and their stomach in place. After measuring for about 5 seconds, the scale showed their height and weight, and then participants left the scale. BMI in Kg/m² was calculated by dividing weight in kilograms by the square of height in meters. Participants stood on a horizontal floor to measure waist and hip circumference with their feet joined together and their arms sagged naturally, spreading their weight evenly across their legs. When measuring waist circumference, participants were told to expose the abdominal skin and relax (belly without clothes), breathing normally, no sucking in their belly or holding their breath. Placing the tape around a horizontal surface about 1cm above the navel so that it closed to the skin but did not press against the skin. When measuring the hip circumference, the tape should be tightly against pants at the widest part of the buttocks and wrapped in the surface but should be avoided getting trapped. The accurate of waist circumference and hip circumference were 0.1cm. During the measurement, participants straightened their back without pressing their abdomen and leaning forward, slightly to the left of their front, extended their arms into the test site and rest their elbows on the elbow pads. When the measurement was ready, press the switch and keep relaxed. Before displaying the results, participants were reminded not to speak, not to move their arms, and to wait for their blood pressure to appear. All the measurement was repeated three times according to the above operation, and the average value of three times was taken to be the final data.

Thyroid scan

The affiniti50 color Doppler ultrasound system from Philips was used to check the thyroid gland with a linear array probe and a frequency of 5-12MHz. Participants took a conventional supine position with their heads back, heads down, and necks high to allow full exposure of the anterior neck area. Continuous top-down cross-section, longitudinal section, and multi-section scans were performed to record the overall thyroid and intrathyroidal nodules in detail. The description of intrathyroidal nodules covered the site, echo, number, size, aspect ratio, margin, border, morphology, stun, internal structure, calcification, blood flow signal, etc. Thyroid nodules were classified as different levels according to Kwak-TIRADS ^[59] criteria.

Fine-needle aspiration (FNA) biopsy

Color Doppler ultrasound is used to examine the thyroid when performing the FNA. The first step was to observe the target nodule and locate it on the body. The patient was in a supine position with the shoulders raised so that the head was overextended and fully exposed to the anterior neck area. After skin disinfected and sterile drapes whisked, the operator fixed the ultrasound probe with one hand, and the other held the 22G puncture needle along the scan section and inserted it obliquely for real-time observation. When the needle tip reached the center of the nodule, stop inserting the needle, pull out the needle core, and pull it back and forth 4-5 times in different needle paths within 5s (liquid-based can puncture about 10 times), quickly withdraw the needle, and press the puncture point with gauze. Patients were advised to avoid swallowing during puncture. The aspirate was pushed onto a glass slide, smeared evenly, and immediately fixed in 95% absolute alcohol. Specimens were sent to the cell room of the pathology department.

DNA extraction, shotgun metagenomic sequencing and reads quantity control

DNA from a total of 1,259 fecal samples was extracted using MagPure Stool DNA KF kit according to the instructions of the manufacturer. DNA nanoball (DNB) based DNA library construction and combinatorial probe-anchor synthesis (cPAS) based shotgun metagenomic sequencing with 100bp paired-end reads were applied to all samples (MGI2000, MGI, Shenzhen, China). Quality control (QC) of raw sequencing reads was applied to filter out low-quality using overall accuracy ($OA \geq 0.8$) control strategy as previously described^[60]. The high-quality reads were aligned to hg19 using SOAPaligner/soap2 to filter out human reads (identity ≥ 0.9) (RRID:SCR_005503). On average 12.5 Gb (± 3.3 Gb) high-quality non-human data was generated per sample. Totally we got 15.73 TB data for downstream analysis.

Microbiota composition profiling

Taxonomic annotation and quantification were performed based on MetaPhlan2 with default settings^[61], generating gut microbial profiling including bacteria, archaea, eukaryotes and viruses. Taxon-specific community functional profiles were further generated using HUMAnN2 (the HMP Unified Metabolic Analysis Network 2).^[62]

Principal coordinates analysis (PCoA)

PCoA were performed based on species or pathway relative abundance by using R program “ade4” package. Four distances were used including Hellinger distance, (JSD), spearman coefficient distance(SCD) and pearson coefficient distance (PCD).

Enterotype analysis

We used analysis flow introduced by Arumugam.M^[63] to identify enterotype number. Combination of four distances (Hellinger, JSD, SCD and PCD) and two clustering methods (complete and k-means) were used,

with different clustering assessment indices from “Numpy” package, finally major-rule was used to identify enterotype number as two.

Microbiota diversity and gini index calculation

We used R program “vegan” package to calculate shannon, simpson index and observed species number (OBS) in each sample. Gini index was calculated with “ineq” package. Gini index of specy composition was calculated based on 459 randomly chosen species, for fair comparison in considering more species were detected than pathways.

OBS comparison based on resampling subsets

40 samples were randomly chosen from TN2 and TN3 respectively. Hypothesis tests of gender ratio (fisher’s exact test), BMI (student’s t test) and age (student’s t test) were performed .Only if no significant differences (p-value >0.4) were detected in all three features, then this subset was used to compare OBS between TN4 and TN2 or Control, finally we used median value of all p-value (Wilcoxon rank sum test) as adjusted p-value.

Association between microbiota and thyroid nodule or thyroid function

MaAsLin2 (Multivariate Association with Linear Models)^[64] was used to associate microbiota and thyroid nodule or functions with default parameters except for “-p 0.2”, while controlling for demographic covariates (gender, age, BMI). Multiple adjusting was built-in BH method. Positive coefficients of microbial species or pathways with HTN group (or TSH, FT4, FT4) were considered with positive associations. Covariates indicate BMI, age and gender unless with otherwise mention in Method section.

Univariate covariates adjusting: R program “glmnet” package and analysis of variance model(AOV) was used to assess correlation between microbiota (including OBS) and covariates with group(TI-RADS or HTN group).

Microbiota inner-interaction assessment

We used spearman correlation coefficient and microbiota relative abundance, to identify microbial inner-interactions, ie, interactions between pathway and specy, specy and specy, pathway and pathway. We randomly subsampling 80% of all samples to bypass possible outliers and repeat this process 100 times, finally used mean spearman correlation coefficient and mean p-value as inner-interaction measurements.

Statistic analysis

Participants’ demographic, lab test and life style characteristics were summarized with a standardized statistical significance test method ^[65], where the test pathway mainly adopts one way ANOVA and chi-square test (**Figure S16**), the results includes control vs nodule, control vs TN2, TN3 vs TN4, LTN

(control+TN2) vs HTN (TN3+TN4) for female and male (**Table S1**). Categorical variables were shown as counts and percentages (n(%), missing n(%)), and associations were tested using a chi-square test. Continuous variables were shown as mean (\pm standard deviation), and differences between groups were analyzed in detail using one way ANOVA and normality test (Shapiro.test), then homogeneity test (Bartlett.test or Levene.test), and finally parameter test (aov.test or kruskal.test) and non-parametric test (oneway.test). Statistical analyses on cohort characteristics were performed on R version 3.6.1. A two-sided p value < 0.05 was considered statistically significant. R program “lars” package was used to assess correlation between covariates with LTN/HTN group and thyroid functions by using LASSO (Least absolute shrinkage and election operator) regression.

Discriminative models construction

We designed a whole process to predict thyroid nodule by merging 1244 participants’ demographic information, species profiles, and pathway (**Figure S14**). This process mainly includes data preprocessing, feature selection, and disease prediction. Specifically, for data preprocessing, data standardization makes all values between 0 and 1. For feature selection, three methods were used to identify the representative features, (1) Filter method: constant variable and variance minimum variable elimination method reduced the original 1392 features to 596 features, strongly correlated variable elimination method further filtered 100 features, and 496 features were preserved. (2) Wrapper method: we trained a Random Forest using tenfold cross-validation to obtain prediction accuracy and adopted backward elimination method to obtain 104 features from the previous feature sets (**Figure S15a**). (3) Embedded method: we split 1244 participants into train data and test data in a ratio of seven to three, used a Gradient Boosting Machine (GBM) with tenfold cross-validation for train data to compute relative influence of each feature, and finally identify 53 features by ranking and accumulating the relative influence up to 85% model contribution (Fig. 4a). For disease prediction model, GBM was used to predict thyroid nodule^[66]. (1) Model training: we trained a GBM using tenfold cross-validation for train data, set the distribution function as “Bernoulli” (logistic regression for 0-1 outcomes) and gridding parameters (e.g., interaction.depth, shrinkage, n.minobsinnode, and bag.fraction) to search the optimal parameter values for minimum error, and the loss function is defined $\sum (y_i - \hat{y}_i)^2$, where y_i is the true value and \hat{y}_i is predictive value. (2) Model testing: we selected top-5 minimum errors and the corresponding parameter values to obtain the highest AUC and accuracy for test data, and finally determine a group of parameter values (shrinkage=0.1, interaction.depth=2, n.minobsinnode=1, bag.fraction=0.8, and n.tree=300) based on the highest AUC (0.734) and accuracy (0.697) (Figure S15b). (3) Independent verification: we firstly used the trained model to measure the AUC (0.739) for all validation data (294 participants) (Fig. 4e), then to obtain robust estimates of model performance, we randomly selected 200 participants from the validation data, and re-run the trained model 100 times to obtain AUC across runs (**Fig. 4g**).

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee from the First Affiliated Hospital of Zhengzhou University and informed consent was obtained from all subjects.

Consent for publication

Not applicable.

Availability of data and materials

All shotgun whole genome sequences have been uploaded into The European Nucleotide Archive PRJEB36271. Source data including microbial species and functional composition, analysis and related code are deposited in <https://github.com/xiaoshubaba/CHIP-thyroid>. Demographic features and serum thyroid functions used in this study are available for only academic usage, please contact Professor Suying Ding (fccdingsy@zzu.edu.cn) for metadata request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was equally funded and supported by Chinese National Science and Technology Major Project 2018ZX10305410, Henan Province Key Science and Technology program 172102310049, Henan province Medical Science and Technique Project 2018020001 grant and Henan province postdoctoral research grant 001801005.

Authors' contributions

D.S.Y., Y.Z.J. and Z.Q. designed the study. D.S.Y. recruited the participants and managed the cohort. L.T.T, W.L., L.J.M, D.Y, H.Z., H.M., Z.L.H., F.Q., H.H.D and Y.X. collected physiological and questionnaire data. G.X.X., Y.H., W.Y., J.M.F., S.Q.M., L.Y., R.Z.G. and R.B.C collected biological samples. Y.D.T., L.H.Q. and M.R.S. comprised the clinical team. L.A. performed the bioinformatics analysis, L.A. and C.J.F. did the statistical analysis and modeling. L.A. and C.J.F. wrote the manuscript. L.A., L.T.T., G.X.X., Y.H., W.Y. and C.J.F. equally contributed to this work.

Acknowledgements

We gratefully thank all participants enrolled in CHIP cohort and nurses or clinicians in The First Affiliated Hospital of Zhengzhou University who assisted with sample and questionnaire collections.

References

1. JK H, WK L, M L, D J, S F: US Features of thyroid malignancy: pearls and pitfalls. Radiographics : a review publication of the Radiological Society of North America, Inc 2007, 27:847-860; discussion

861-845.

2. C D, G C, G L, R B, D M, A P, E P, M T, S T, M A, et al: The natural history of benign thyroid nodules. *JAMA* 2015, 313:926-935.
3. DS C, GM D, BR H, BR H, RT K, SL L, SJ M, EL M, B M, F P, et al: Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid : official journal of the American Thyroid Association* 2009, 19:1167-1214.
4. E O, G W, KM W, J D, PT S: Epidemiology of Thyroid Cancer: A Review of the National Cancer Database, 2000-2013. *Cureus* 2019, 11:e4127.
5. Y Y, X C, S W, L G, H Z, Q Z, J T, F L, Y Z, F L, Y H: Thyroid nodules in centenarians: prevalence and relationship to lifestyle characteristics and dietary habits. *Clinical interventions in aging* 2018, 13:515-522.
6. D D, J P, J K, R V, J V-K, A S, J C, V S, L D-P, A N, B Z: Factors Associated with the Prevalence of Thyroid Nodules and Goiter in Middle-Aged Euthyroid Subjects. *International journal of endocrinology* 2017, 2017:8401518.
7. Y H, EK L, HY S, SW I, SJ K, JW Y, MJ K, J K, HS C, CH R, et al: Genome-Wide Association Study Reveals Distinct Genetic Susceptibility of Thyroid Nodules From Thyroid Cancer. *The Journal of clinical endocrinology and metabolism* 2018, 103:4384-4394.
8. Sarne D: Effects of the Environment, Chemicals and Drugs on Thyroid Function. In *Endotext*. Edited by Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, Hershman JM, Kaltsas G, Koch C, Kopp P, et al. South Dartmouth (MA); 2000
9. PN T, E P, S C, PJ C, M T, SJ B, BH M, HA S, J M, K W, et al: Whole-genome sequence-based analysis of thyroid function. *Nature communications* 2015, 6:5681.
10. E F, R W: Microbiota and Thyroid Interaction in Health and Disease. *Trends in endocrinology and metabolism: TEM* 2019, 30:479-490.
11. J Z, F Z, C Z, Q X, C L, Y Y, H W, Y S, Y W, X M, et al: Dysbiosis of the gut microbiome is associated with thyroid cancer and thyroid nodules and correlated with clinical index of thyroid function. *Endocrine* 2019, 64:564-574.
12. MR M, AD W, BJ C: Consistent and correctable bias in metagenomic sequencing experiments. *eLife* 2019, 8:undefined.
13. GF M, MR G, PM F, A R, FA CJ, DR N, RG G: Fine needle aspiration biopsy of thyroid nodule smaller than 1.0 cm: accuracy of TIRADS classification system in more than 1000 nodules. *The British journal of radiology* 2018, 91:20170642.
14. D R-C, N S, M G, CG dLR-G: Bacteroides Shaping the Metabolism of Intestinal Population through Diet to Improve Human Health. *Frontiers in microbiology* 2017, 8:376.
15. EP N, CH D, SS R: The role of microbial amino acid metabolism in host metabolism. *Nutrients* 2015, 7:2930-2946.

16. HL Z, Y C, XL X, YX Y: Effects of Branched-chain Amino Acids on In vitro Ruminant Fermentation of Wheat Straw. *Asian-Australasian journal of animal sciences* 2013, 26:523-528.
17. FQ H, J L, L J, FX W, RN A, MJ W, WJ M, G M, YJ Z, SL W, et al: Serum-plasma matched metabolomics for comprehensive characterization of benign thyroid nodule and papillary thyroid carcinoma. *International journal of cancer* 2019, 144:868-876.
18. Ryoo I, Kwon H, Kim SC, Jung SC, Yeom JA, Shin HS, Cho HR, Yun TJ, Choi SH, Sohn CH, et al: Metabolomic analysis of percutaneous fine-needle aspiration specimens of thyroid nodules: Potential application for the preoperative diagnosis of thyroid cancer. *Sci Rep* 2016, 6:30075.
19. RLE G, EMMS G: Metabolomics of thyroid nodules and the future. *Archives of endocrinology and metabolism* 2018, 62:493-494.
20. A K, F DV, P K-D, F B: From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* 2016, 165:1332-1345.
21. A B, SH D, G H, SE A, GE L, HJ F: Impact of pH on lactate formation and utilization by human fecal microbial communities. *Applied and environmental microbiology* 2007, 73:6526-6533.
22. Y T, X H, X Z, H J, Z X, Q T, X L: Production of Butyrate from Lactate by a Newly Isolated *Clostridium* sp. BPY5. *Applied biochemistry and biotechnology* 2016, 179:361-374.
23. C B, RJ R, L M, F K, E T, C C, C M: Lactate is mainly fermented to butyrate by human intestinal microfloras but inter-individual variation is evident. *Journal of applied microbiology* 2005, 99:201-212.
24. JR D: Inhibition of histone deacetylase activity by butyrate. *The Journal of nutrition* 2003, 133:2485S-2493S.
25. K S, MS B, SP P, J S, DV F: Butyrate histone deacetylase inhibitors. *BioResearch open access* 2012, 1:192-198.
26. Y T, J N, S H, F R, Y U, T M, H K, Y F, K Y, K K, N W: Sodium butyrate induces growth arrest and senescence-like phenotypes in gynecologic cancer cells. *International journal of cancer* 2001, 94:257-267.
27. M R-P, M E: Cancer epigenetics reaches mainstream oncology. *Nature medicine* 2011, 17:330-339.
28. M P, L T, L R, C R, MA P, C R, GA C, MA D, GJ J: Radiosensitivity enhancement of human thyroid carcinoma cells by the inhibitors of histone deacetylase sodium butyrate and valproic acid. *Molecular and cellular endocrinology* 2018, 478:141-150.
29. CA B, WS G: *Fusobacterium nucleatum* - symbiont, opportunist and oncobacterium. *Nature reviews Microbiology* 2019, 17:156-166.
30. JH H, G C, T B, W H, M C, G M: Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* 2010, 464:908-912.
31. JH H, AG K, NA P, EC M, AB B: Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109:19786-19791.

32. MJ P, MP F, K K, FE D: Increased iodine uptake in thyroid carcinoma after treatment with sodium butyrate and decitabine (5-Aza-dC). *Otolaryngology–head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 2007, 137:722-728.
33. MA L: Sodium butyrate selectively alters thyroid hormone receptor gene expression in GH3 cells. *The Journal of biological chemistry* 1990, 265:17474-17477.
34. J O-C, F M, A P, A A: Modulation of thyroid hormone nuclear receptors by short-chain fatty acids in glial C6 cells. Role of histone acetylation. *The Journal of biological chemistry* 1986, 261:13997-14004.
35. C H, OA R-Z, JF A-D, F G-D, P P-M, J D-L, GM Q-N, BB L, JA N-C, M T-S, et al: Intestinal Microbiota Is Influenced by Gender and Body Mass Index. *PloS one* 2016, 11:e0154090.
36. SC D, JS Y, SD B, BK R: Gut microbiome diversity influenced more by the Westernized dietary regime than the body mass index as assessed using effect size statistic. *MicrobiologyOpen* 2017, 6:undefined.
37. P C, S Y, CR S, PJ A, EM M, RP R, C S: Prebiotics from Seaweeds: An Ocean of Opportunity? *Marine drugs* 2019, 17:undefined.
38. W W, L C, R Z, X W, L S, S H, G W, B X: Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *Journal of clinical microbiology* 2014, 52:398-406.
39. SA S, GP V, ER G, DA M, AG V: TRH protection against memory retrieval deficits is independent of endocrine effects. *Pharmacology, biochemistry, and behavior* 1992, 41:145-152.
40. E F, G I, J C, R N, M N: Insulin resistance and normal thyroid hormone levels: prospective study and metabolomic analysis. *American journal of physiology Endocrinology and metabolism* 2017, 312:E429-E436.
41. SI S: Tyrosine kinase inhibitors and the thyroid. *Best practice & research Clinical endocrinology & metabolism* 2009, 23:713-722.
42. Khaliq W, Andreis DT, Kleyman A, Gräler M, Singer M: Reductions in tyrosine levels are associated with thyroid hormone and catecholamine disturbances in sepsis. *Intensive Care Medicine Experimental* 2015, 3.
43. B D, C L: Method for measuring ATP production in isolated mitochondria: ATP production in brain and liver mitochondria of Fischer-344 rats with age and caloric restriction. *American journal of physiology Regulatory, integrative and comparative physiology* 2003, 285:R1259-1267.
44. G B, R M, S R, C M, S C, A O, L E, C L, A A, A P, et al: Fasting induces anti-Warburg effect that increases respiration but reduces ATP-synthesis to promote apoptosis in colon cancer models. *Oncotarget* 2015, 6:11806-11819.
45. YP L, HH H: Effects of thyroid hormones on amino acid and protein metabolism–IV. Effects of thyroid hormone on amino acid concentration and incorporation of [¹⁴C]L-valine in vivo and in vitro into proteins of the guinea-pig. *The International journal of biochemistry* 1982, 14:295-304.

46. AN B, MS C-R, A B, D P, FK R, ML B-C, JJ S: Thyroid hormones alter the adenine nucleotide hydrolysis in adult rat blood serum. *BioFactors* (Oxford, England) 2011, 37:40-45.
47. A A, S A, D M, I S: Effect of High-dose Biotin on Thyroid Function Tests: Case Report and Literature Review. *Cureus* 2018, 10:e2845.
48. BM K, AJ L, LJ D, AS J, NA B: Prevalence of biotin supplement usage in outpatients and plasma biotin concentrations in patients presenting to the emergency department. *Clinical biochemistry* 2018, 60:11-16.
49. SA O, AA M, NS H: How Biotin Induces Misleading Results in Thyroid Bioassays: Case Series. *Cureus* 2019, 11:e4727.
50. C F, M C, G B, A A, A S, F A, G P, A B, G C: Ultrasound imaging classifications of thyroid nodules for malignancy risk stratification and clinical management: state of the art. *Gland surgery* 2019, 8:S233-S244.
51. AM V, J W, E S, TD S: Role of the gut microbiota in nutrition and health. *BMJ (Clinical research ed)* 2018, 361:k2179.
52. CL G, TL W: The gut microbiota at the intersection of diet and human health. *Science (New York, NY)* 2018, 362:776-780.
53. AM M, BM A-A, BM A-R: Vitamin d deficiency and its association with thyroid disease. *International journal of health sciences* 2013, 7:267-275.
54. Prashant Tripathi NS, Manish Kumar Verma, Anand Narayan Singh: Association of Vitamin B12, Folate and Ferritin with Thyroid Hormones in Hypothyroidism. *Annals of International Medical and Dental Research* 2019, Vol (5), Issue (1).
55. M S, T L, D AK, A R, M B, AM L-Č, H G, D R: Factors Affecting Gastrointestinal Absorption of Levothyroxine: A Review. *Clinical therapeutics* 2017, 39:378-403.
56. K Y, K H, K S, J K: Metabolism of Dietary and Microbial Vitamin B Family in the Regulation of Host Immunity. *Frontiers in nutrition* 2019, 6:48.
57. P D, P B, J N: Metagenomic analysis of microbe-mediated vitamin metabolism in the human gut microbiome. *BMC genomics* 2019, 20:208.
58. KS F, M W, M N, RG I: Potential of butyrate to influence food intake in mice and men. *Gut* 2018, 67:1203-1204.
59. JY K, KH H, JH Y, HJ M, EJ S, SH P, HK J, JS C, BM K, EK K: Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology* 2011, 260:892-899.
60. C F, H Z, Y L, B C, M H, H R, H L, JM L, M X, W L, et al: Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *GigaScience* 2018, 7:1-8.
61. DT T, EA F, TL T, M S, G W, E P, A T, C H, N S: MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature methods* 2015, 12:902-903.

62. EA F, LJ M, G R, LR T, M S, G W, KS L, R K, JG C, N S, C H: Species-level functional profiling of metagenomes and metatranscriptomes. *Nature methods* 2018, 15:962-968.
63. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, et al: Enterotypes of the human gut microbiome. *Nature* 2011, 473:174-180.
64. Multivariable Association in Population-scale Meta'omic Surveys
65. Hewson P: *Statistical Rethinking: a Bayesian Course with Examples in R and Stan*. R. McElreath, 2015 Boca Raton Chapman and Hall–CRC 470 pp., £60.99 ISBN 978-1-482-25344-3. *Journal of the Royal Statistical Society* 2016, 179:1131-1132.
66. Friedman JH: *Greedy Function Approximation: A Gradient Boosting Machine*. *The Annals of Statistics* 2001.

Figures

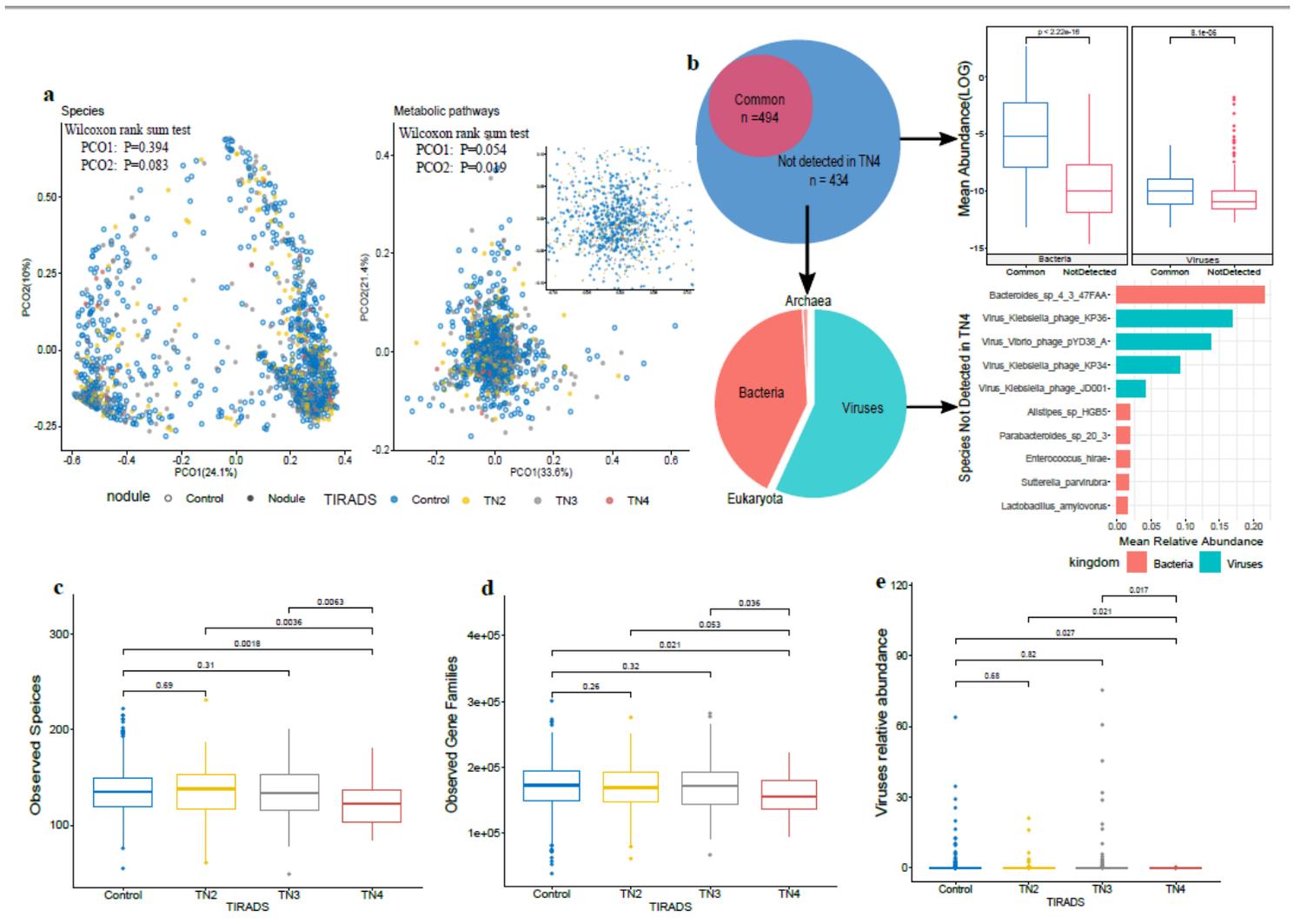


Figure 1

Microbial species distributions. Significance determined by single sided wilcoxon's rank sum test. For all boxplots in this study, center is median, boxes interquartile range (IQR), whisker is 1.5×IQR. a, Principal coordinate analysis of spearman coefficient distance, based on relative abundance of species (left panel) and pathways (right panel). Colored dots represented samples with TI-RADS grade, x-axis was PC01 and y-axis was PC02. b, Venn diagram (top left) indicated shared species in all groups (n=494) and species not detected in TN4 (n=434); Kingdom assignments of species not detected in TN4 was illustrated in pie plot (bottom left); boxplots showed distribution of mean relative abundance of Bacteria and Virus species detected in all group and not detected in TN4 (top right); Bar plots indicated top 10 species not detected in TN4 with mean relative abundance as bar length (bottom right). c, d, e, boxplots showed distribution of microbial characteristics. c, Number of observed species (OBS). d, Number of observed gene families. e, Viruses relative abundance.

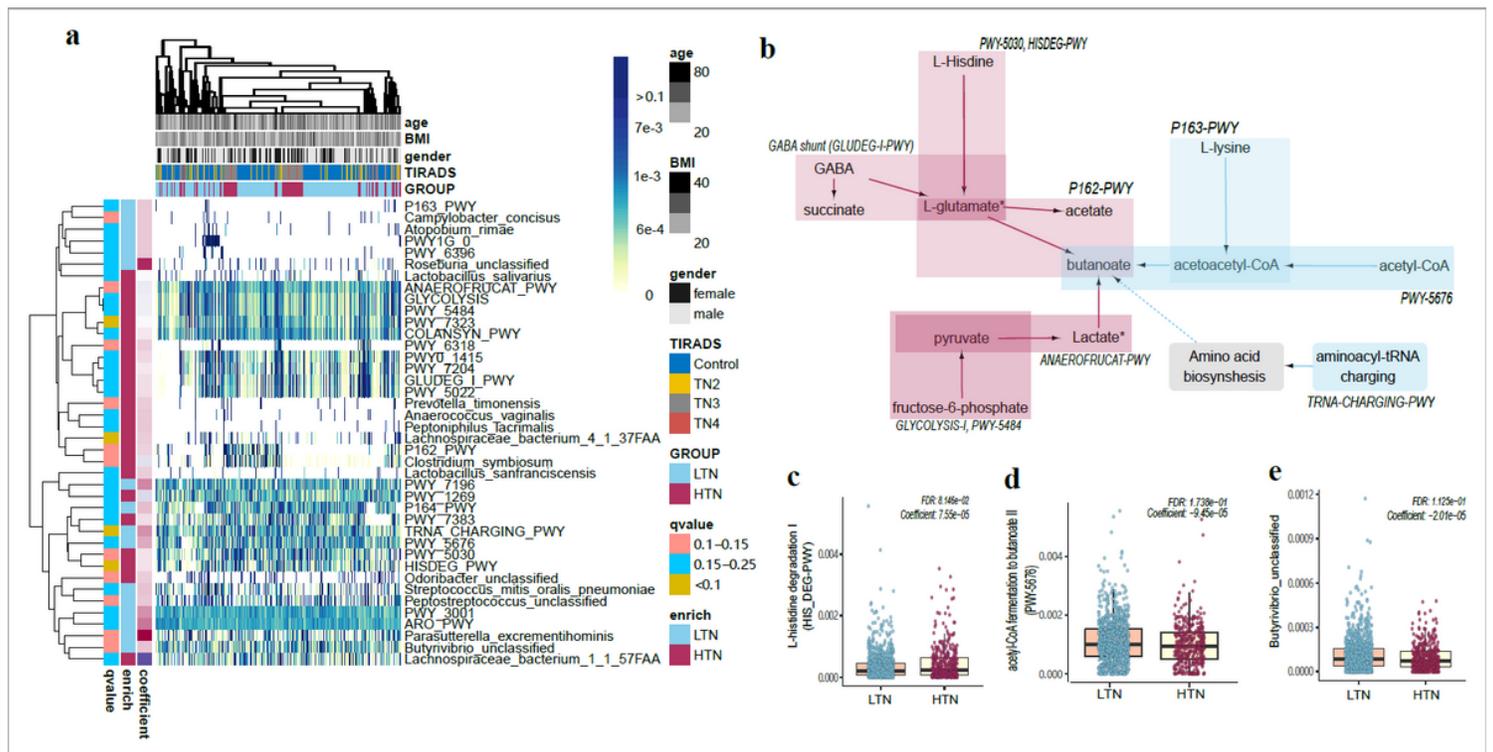


Figure 2

Microbiota associated with high grade thyroid nodule (HTN). a, Heatmap of statistically significant (see method) HTN associated microbiota species and pathways in all samples (n=1259). Gradient color in heatmap indicated relative abundance (standardization by rows), phylogenetic trees were generated by using complete clustering method based on binary distances. Enrollment age, BMI, gender, TI-RADS level and LTN/HTN group were annotated with color bars attached to column phylogenetic tree. Row color bars represented qvalue category (<0.1, 0.1-0.15, 0.15-0.25), enrichment and coefficient of microbial species/functions with HTN group. b, Schematic diagram showing metabolic pathways associated with HTN. Red background in rectangle denoted enriched functions in HTN; blue denoted functions enriched in LTN; grey denoted neutral functions. Transparency of pathway are proportional to its relative abundance. Dashed line arrow and metabolites with star pointed to the inferences that were previously reported. c,

Boxplots showed distribution of L-histidine degradation function, butanoate formation function and Butyrivibrio unclassified in LTN and HTN, significance determined in multiple linear regression and FDR adjust with Masslin2.

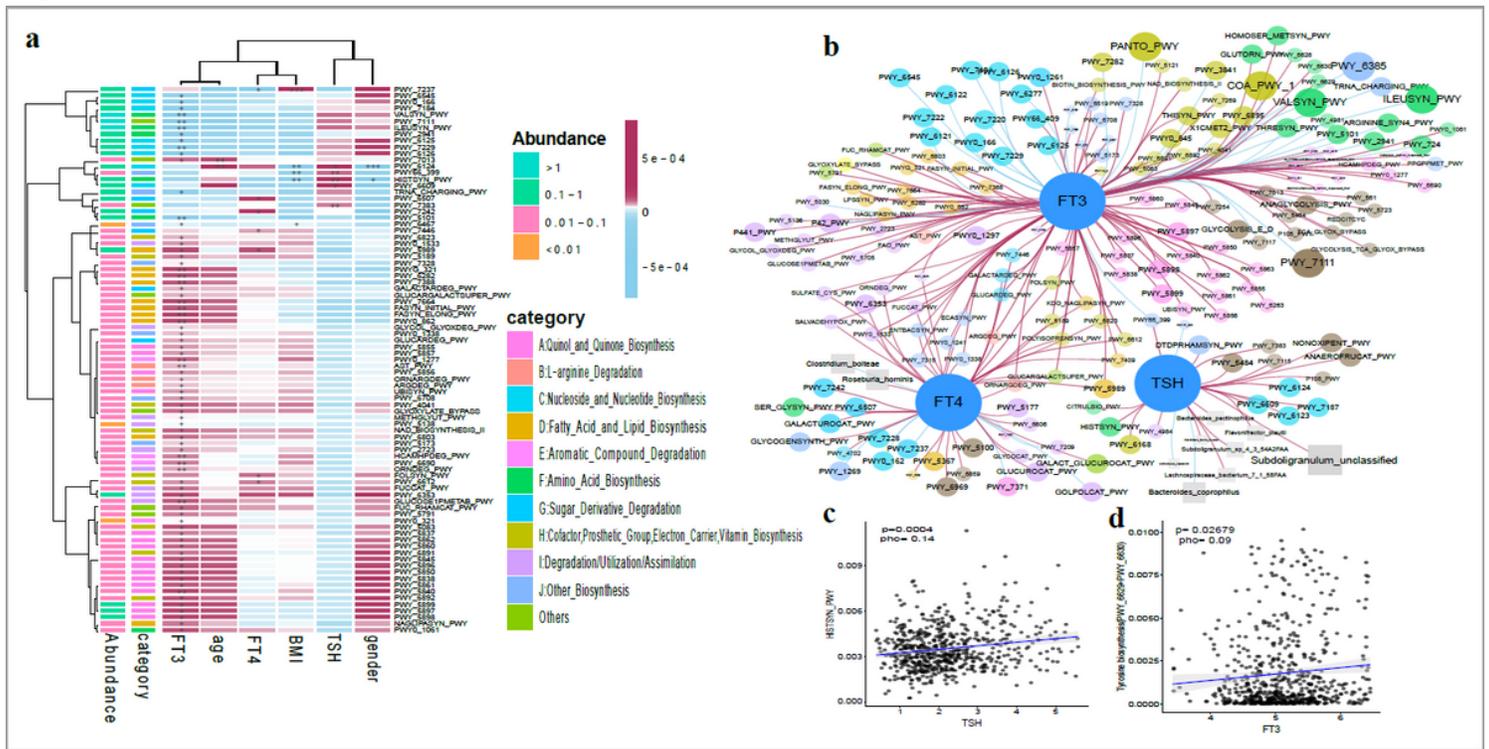


Figure 3

Microbiota associated with thyroid functions. a, Heatmap of statistically significant (Method) correlations between thyroid functions and microbiota in all participants (n=1259). Gradient color in heatmap indicated coefficient with metadata; significance was denoted in cells as follows: +++, qvalue < 0.01; ++, qvalue 0.01-0.1; +, qvalue 0.1-0.25. Row color bars represented relative abundance (<0.01, 0.01-0.1, 0.1-1, >1) and function category. b, Significant associations among pathways or species and thyroid functions. Node size and line width were proportional to species relative abundance and spearman coefficient. Node color indicated pathway function category (the same as in Figure 3a), line color indicated positive (red) or negative (blue) association. Thyroid functions (TSH, FT3 and FT4) were emphasized with blue circle. c, d, Scatter plots showing correlation between amino acid biosynthesis and thyroid functions, significance determined by associate test with spearman's rho estimate. c, L-histidine biosynthesis and TSH. d, tyrosine biosynthesis and FT3, y-axis is sum of relative abundance in PWY-6629 and PWY-6630.

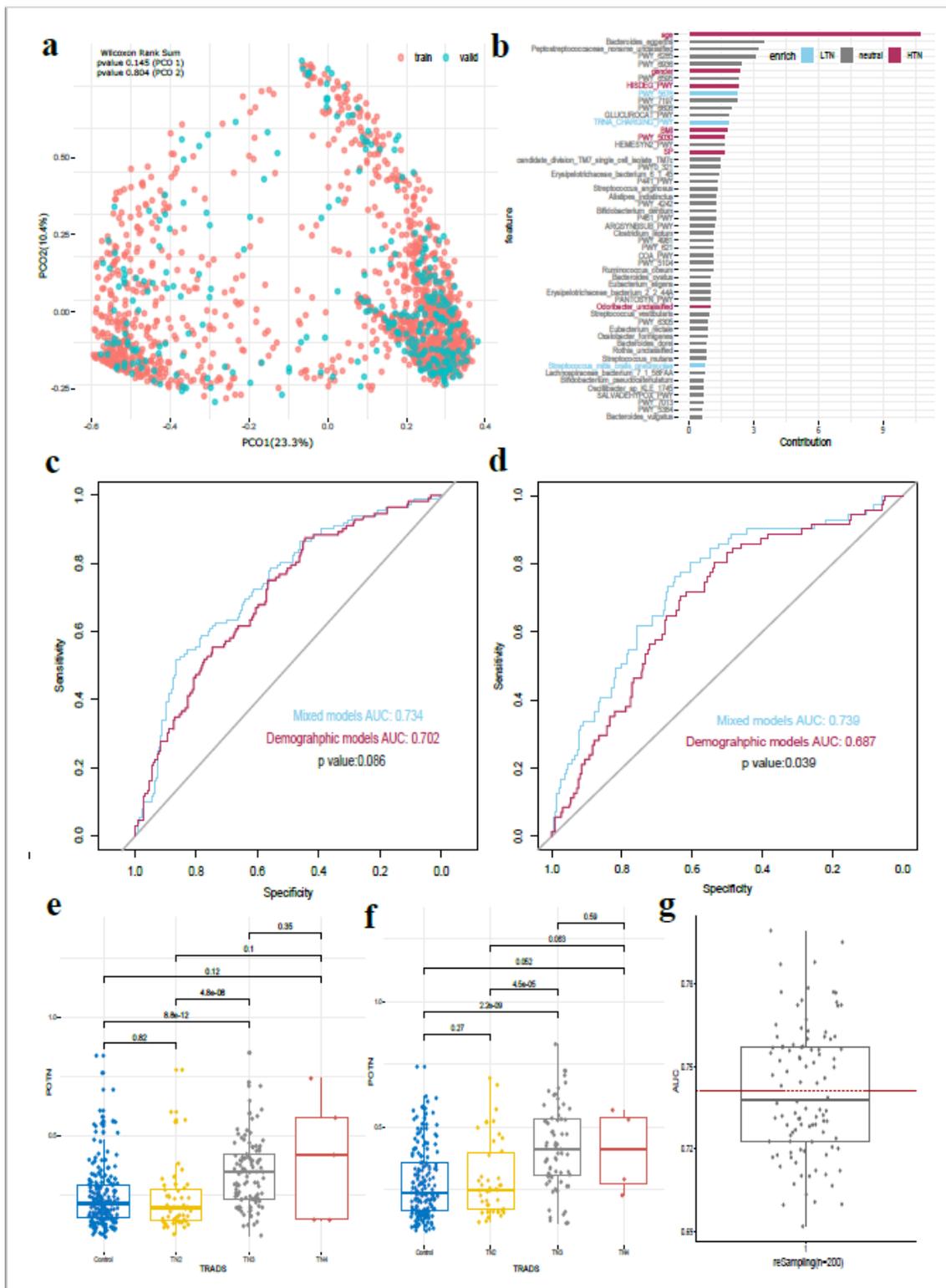


Figure 4

Models to distinguish advanced TIRADS thyroid nodules. a, Principal coordinate analysis of Spearman coefficient distance based on relative abundance of species. Colored dots represented train (n=1259) and validate (n=294) samples, significance determined by two-sided Wilcoxon's rank sum test. b, Barplot indicates 53 features to construct classification models, bar length is proportional to feature relative influence, colored bars indicate features enriched in LTN (blue), HTN (red) or without group significance (grey). c, d

Performance of the classifiers using AUCs, significance determined by single sided AUC comparison by using bootstrap method with 10000 permutations (boot.n=10000). (c) AUC in 300 cross validate samples. (d) AUC in 294 validate samples. e,f Boxplots showing distribution of probabilities of TN (POTN) among different TI-RADS grade, significance in boxplots determined by single sided wilcox's rank sum test. (e) POTN in cross validate samples. (f) POTN in validate samples. g. Boxplot showing distribution of generalized AUCs based on bootstrapped resampling, dash line indicated mean generalized AUC.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementrayinformation.docx](#)
- [SupplementaryTables.xlsx](#)