SUPPLEMENTAL DATA

SUPPLEMENTARY MATERIALS AND METHODS

Bioinformatics analysis

Bioinformatics analysis of miRNAs was carried out by Sistemas Genómicos (Valencia, Spain) as previously reported [1]. For mRNA analysis, quality of data was analyzed using the fastQC program [2]. Reads were mapped against the hg19 Homo sapiens Genome, using the Tophat2 v2.1.0 algorithm [3]. High quality reads were selected using a Phred Score threshold of 20. Gene quantification was carried out using the htseq_count 0.6.1p1 method [4]. For differential expression between samples, the DESeq2 algorithm was applied [5], and a FoldChange threshold of 2 and pValue adjusted by False Discovery Rate (FDR) of 0.05 were selected.

Hierarchical clustering

Hierarchical clustering was performed using gplots package of R software. The average linkage and median centring were the parameters chosen for the unsupervised analysis of differential gene and miRNAs expression.

Functional analysis

Functional enrichment analysis of differential gene expression was performed using clusterProfiler v3.5 [6], with a threshold pValue adjusted by FDR of 0.05. We selected all statistically significant categories present in GO [7] and KEGG database [8].

miRNA-mRNA enrichment analysis and negative correlation

Putative target gene sets of differentially expressed miRNAs were predicted using the following algorithms: MiRanda12, TargetScan14. The p-values were calculated using right-sided

hypergeometric tests, and Benjamini-Hochberg adjustment was used for multiple test correction. An adjusted p-value < 0.05 indicated a statistically significant deviation from the expected distribution. Pearson correlation coefficients between a particular miRNA and its predicted target mRNAs were computed using R (http://www.R-project.org) to determine whether the expression levels of each miRNA and of its mRNA targets were negatively correlated (correlation <0 and Fisher P corrected by FDR <0.05).

Network analysis was performed using Cytoscape v3.5 using the significant Person correlation between mRNAs and miRNAs obtained in the previous step. Cluster identification of the network was performed using MCODE algorithm [9]. Functional analysis of the different network clusters were carried out using BINGO method [10].

REFERENCES

[1] Martínez-Hernández R, Sampedro-Núñez M, Serrano-Somavilla A, Ramos-Leví AM, de la Fuente H, Triviño JC, et al. A MicroRNA Signature for Evaluation of Risk and Severity of Autoimmune Thyroid Diseases. J Clin Endocrinol Metab 2018;103:1139–50. doi:10.1210/jc.2017-02318.

[2] Andrews, S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data n.d. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[3] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009;25:1105–11. doi:10.1093/bioinformatics/btp120.

[4] Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31:166–9. doi:10.1093/bioinformatics/btu638.

[5] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550. doi:10.1186/s13059-014-0550-8.

[6] Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7. doi:10.1089/omi.2011.0118.

[7] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9. doi:10.1038/75556.

[8] Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 2000;28:27–30. doi:10.1093/nar/28.1.27.

[9] Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003;4:2.

[10] Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005;21:3448–9. doi:10.1093/bioinformatics/bti551.

Supplementary Table 1: Primer sequence details of the analyzed genes (a) and NormFinder stabilities values of housekeeping genes (b).

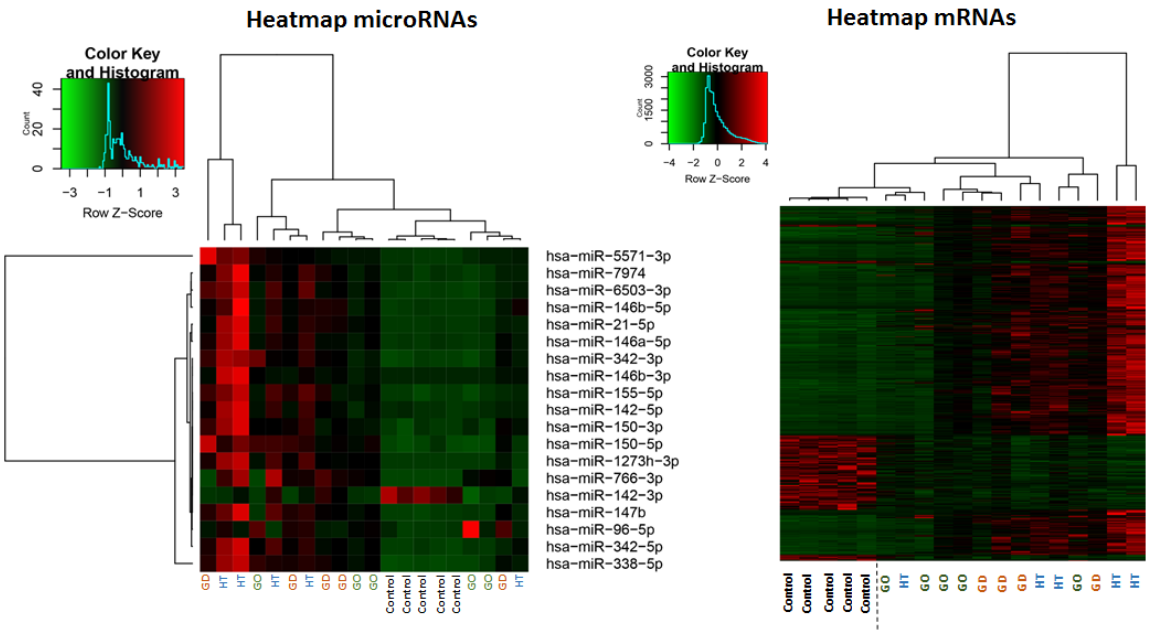
a)

Gene name	Sequence		Product Length
Homo sapiens kinesin family member 27 (KIF27),	FW	CGGCAACTAGCACCATCAG	Amplicon Size = 60
	RV	TCCTTCTGGCTTCAGGACTC	
Homo sapiens enolase 4 (ENO4),	FW	ACAGCAGATCACTGGCAAGA	Amplicon Size = 104
	RV	CCAGGTTGGCACAGATTTCC	
Homo sapiens serine/threonine kinase 36 (STK36)	FW	CAAGGGGTGACCTTTGACC	Amplicon Size = 92
	RV	GGTGGAGTCAACCGAACCT	
Homo sapiens inturned planar cell polarity protein (INTU)	FW	TCAGCGACTCGGGTTCAT	Amplicon Size = 65
	RV	CAGCCATTCAGGCTCAAGA	
Homo sapiens parkin coregulated (PACRG),	FW	AACCGACAGGTCATCTGTGT	Amplicon Size = 92
	RV	CGGTAATAAGGCACCAAGGC	
Homo sapiens Hypoxanthine phosphoribosyltransferase (HPRT)	FW	TGACCTTGATTTATTTTGCATACC	Amplicon Size = 102
	RV	CGAGCAAGACGTTCAGTCCT	
Homo sapiens glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	FW	AGCCACATCGCTCAGACAC	Amplicon Size = 66
	RV	GCCCAATACGACCAAATCC	
Homo sapiens actin beta(β-ACTIN)	FW	CGGAGTACTTGCGCTCAGGA	Amplicon Size = 88
	RV	GCCGACAGGATGCAGAAGGA	

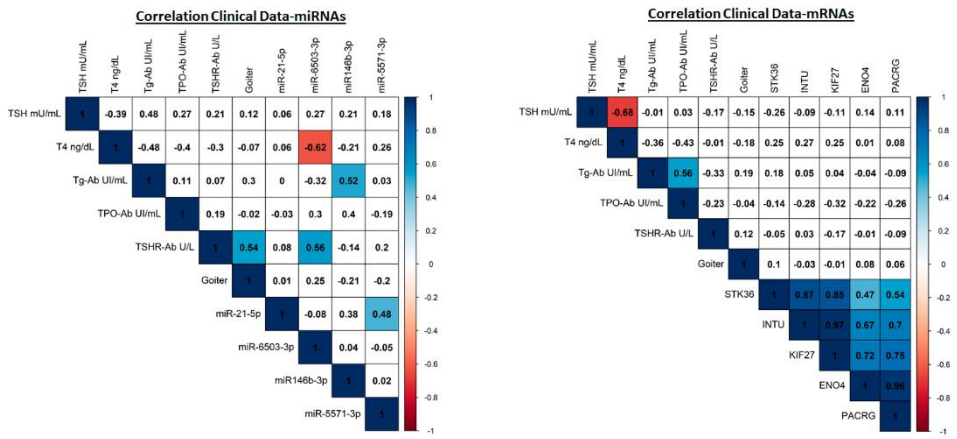
b)

Gene name	Stability value
β-Actin	0.016
GAPDH	0.017
HPRT	0.024
Best Combination Genes β-Actin and GAPDH	0.009

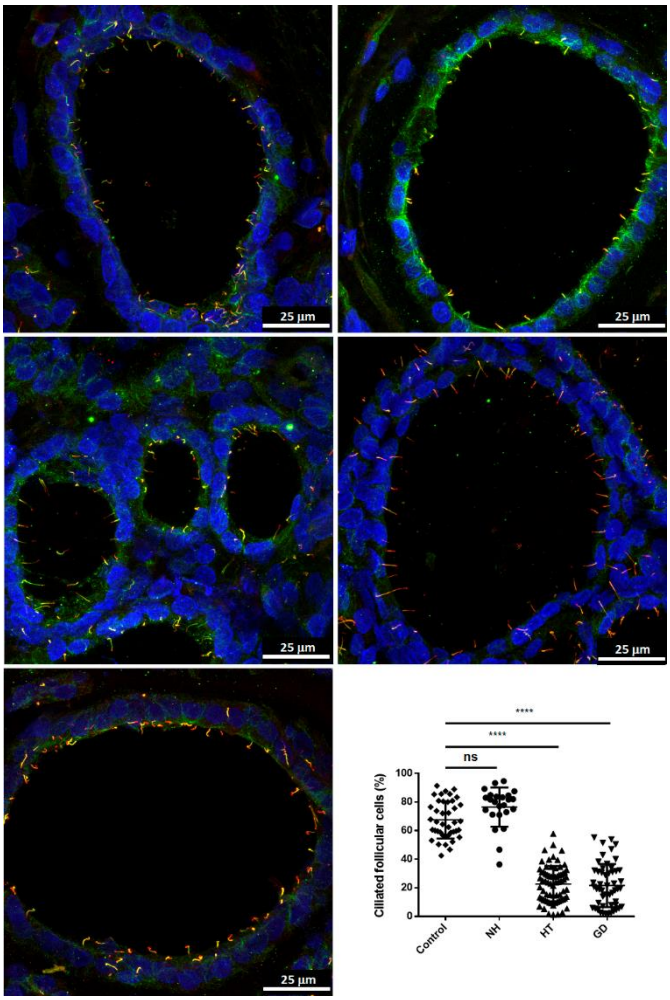
SUPPLEMENTARY FIGURES



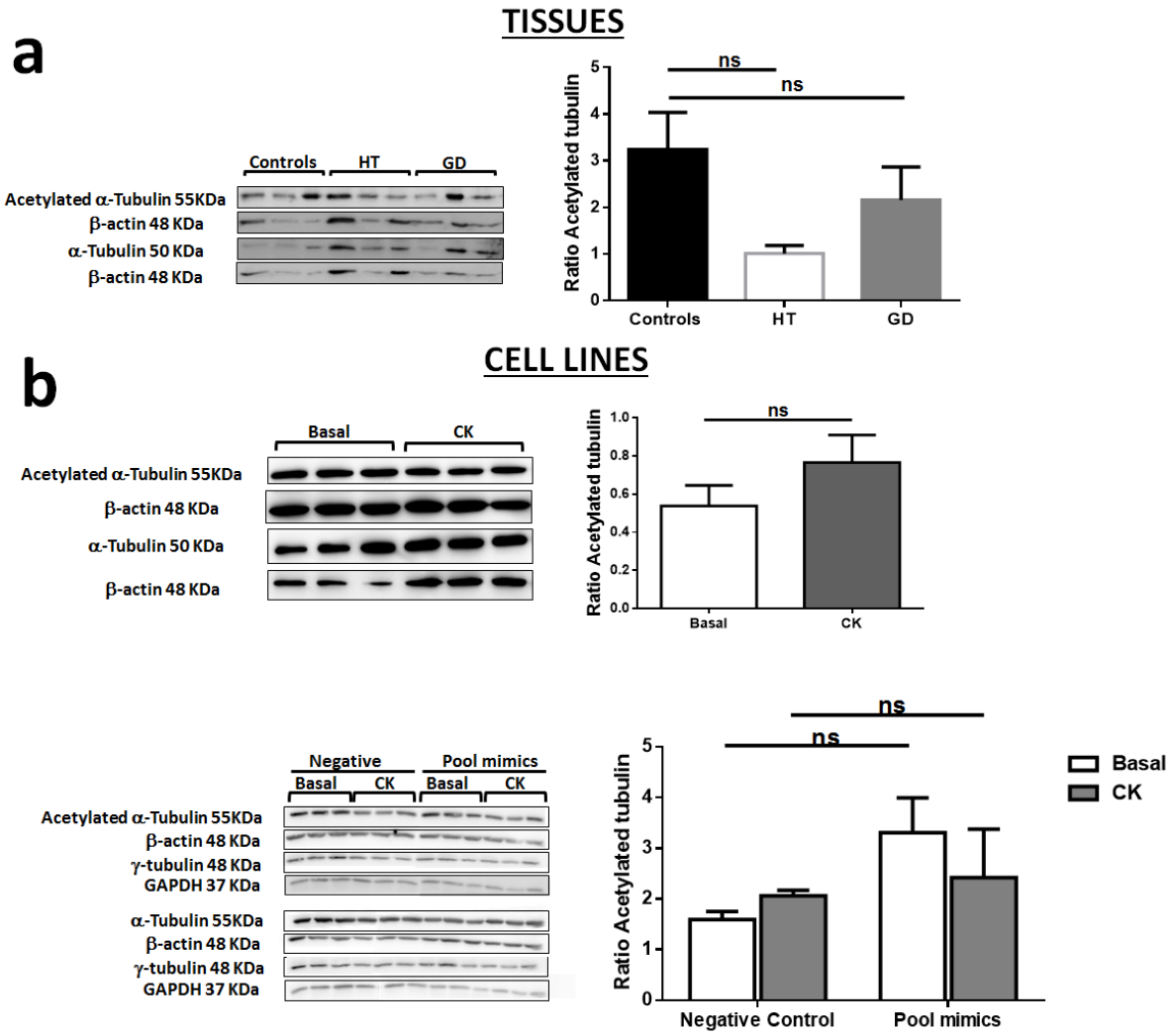
Supplementary Figure 1: Heat map hierarchical clustering of differentially expressed (DE) microRNAs and mRNAs in AITD tissue samples. The graph shows expression in thyroid tissue from AITD samples compared to controls, expressed in fold change. The key color bar indicates that expression levels increased from green to red, depending on the fold change.



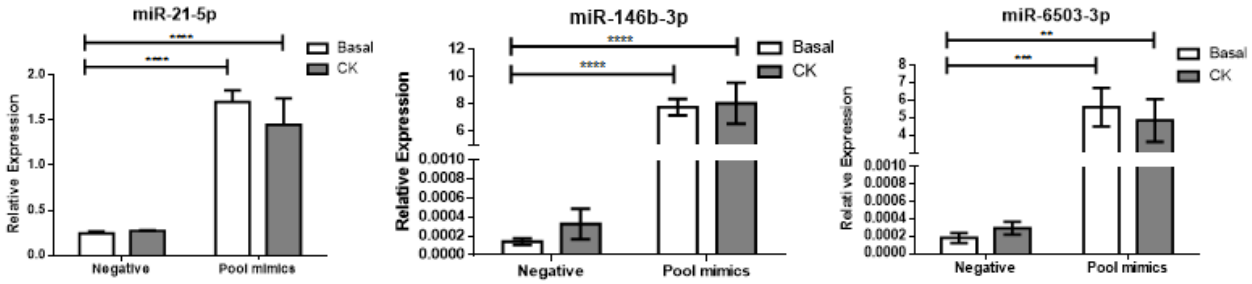
Supplementary Figure 2: Correlation Heatmaps between laboratory parameters of patients and miRNA and mRNA expression levels (Spearman r analysis). Significant negative correlations are shown in red and significant positive correlations in blue. miR-6503-3p was negatively associated with free T4 levels (r=-0.62) and positively correlated with levels of TSHR-Ab (r=0.56). miR-146b-3p were positively associated with Tg-Ab levels (r=0.52). Gene expression of STK36, INTU, KIF27, ENO4 and PACRG did not correlate significantly with clinical parameters.



Supplementary Figure 3: Representative images of five cases of nodular hyperplasia (NH). Immunofluorescence staining of acetylated α-tubulin (green) and Arl13b (red) in thyroid follicles. Quantification of the frequency of follicular ciliated cells showed no statistical significances beteween NH and controls. Data are represented as mean ± SD (****p<0.001 and ns=not significant, by Kruskal-Wallis test).



Supplementary Figure 4: Western blot and quantification of acetylated α-tubulin in thyroid tissue (a) and transfected cell lines (b) protein extractions. No significantly differences were observed between controls and patients nor in transfected cell lines.



Supplementary Figure 5: Relative expression of transfected miRNAs (miR-21-5p, miR-146b-3p and miR-6503-3p) in cell cultures at different conditions (Basal and CK)