

RESEARCH ARTICLE

Gene expression study and pathway analysis of histological subtypes of intestinal metaplasia that progress to gastric cancer

Osmel Companioni^{1*}, José Miguel Sanz-Anquela², María Luisa Pardo³, Eulàlia Puigdecamet⁴, Lara Nonell⁴, Nadia García¹, Verónica Parra Blanco⁵, Consuelo López⁶, Victoria Andreu⁷, Miriam Cuatrecasas⁸, Maddi Garmendia⁹, Javier P. Gisbert¹⁰, Carlos A. Gonzalez¹, Núria Sala^{1*}



1 Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Institut Català d'Oncologia, Barcelona, Spain, **2** Cancer Registry and Pathology Department, Hospital Universitario Príncipe de Asturias and Department of Medicine and Medical Specialties, Faculty of Medicine, University of Alcalá, Alcalá de Henares, Madrid, Spain, **3** Pathology Department, Complejo Hospitalario de Soria, Spain, **4** Microarray Analysis Service, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain, **5** Department of Histopathology, Hospital Universitario Gregorio Marañón, Madrid, Spain, **6** Department of Pathology, Hospital Universitario de la Princesa, Madrid, Spain, **7** Department of Gastroenterology, Hospital de Viladecans, Spain, **8** Department of Pathology, Hospital Clínic de Barcelona, Universitat de Barcelona, Biobanc Clínic IDIBAPS, Barcelona, Spain, **9** Department of Pathology, and Department of Gastroenterology, Hospital Donostia/Instituto Biodonostia, Universidad del País Vasco (UPV/EHU), CIBEREHD, San Sebastián, Spain, **10** Gastroenterology Unit, Hospital Universitario de La Princesa and Instituto de Investigación Sanitaria Princesa (IIS-IP), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Madrid, Spain

* nsala@iconcologia.net (NS); osmel.companioni@yahoo.es (OC)

OPEN ACCESS

Citation: Companioni O, Sanz-Anquela JM, Pardo ML, Puigdecamet E, Nonell L, García N, et al. (2017) Gene expression study and pathway analysis of histological subtypes of intestinal metaplasia that progress to gastric cancer. PLoS ONE 12(4): e0176043. <https://doi.org/10.1371/journal.pone.0176043>

Editor: Ken Mills, Queen's University Belfast, UNITED KINGDOM

Received: September 7, 2016

Accepted: April 4, 2017

Published: April 25, 2017

Copyright: © 2017 Companioni et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Expression data are accessible in Gene Expression Omnibus (GEO): GSE78523. Other relevant data are within the paper and its Supporting Information files.

Funding: Fondo de investigaciones Sanitarias del Instituto de Salud Carlos III, Spanish Ministry of Health. Cofounded by European Regional Development Funds (ERDF/FEDER) "A way to build Europe". <http://www.isciii.es/>. Grant numbers: PI03/0077 (CAG) PI10/01089 (CAG), PI10/01031

Abstract

Background

Intestinal metaplasia (IM) is a precursor lesion that precedes gastric cancer (GC). There are two IM histological subtypes, complete (CIM) and incomplete (IIM), the latter having higher progression rates to GC. This study was aimed at analysing gene expression and molecular processes involved in the progression from normal mucosa to IM, and also from IM subtypes to GC.

Methodology

We used expression data to compare the transcriptome of healthy gastric mucosa to that of IM not progressing to GC, and the transcriptome of IM subtypes that had progressed to GC to those that did not progress. Some deregulated genes were validated and pathway analyses were performed.

Results

Comparison of IM subtypes that had progressed to GC with those that did not progress showed smaller differences in the expression profiles than the comparison of IM that did not progress with healthy mucosa. New transcripts identified in IM not progressing to GC included TRIM, TMEM, homeobox and transporter genes and SNORD116. Comparison to normal mucosa identified non tumoral Warburg effect and melatonin degradation as

(JG), PI10/01203 (JMS) RETICC RD12/0036/0018 (CAG, NS, OC, NG). Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Catalunya, Spain. <http://agaur.gencat.cat/en/inici/index.html>. Grant number: exp. 2014 SGR 726 (CAG, OC, NG, NS). Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), <http://www.idibell.cat/en>, PhD scholarship to OC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

previously unreported processes involved in IM. Overexpressed antigen processing is common to both IM-subtypes progressing to GC, but IIM showed more over-expressed oncogenic genes and molecular processes than CIM.

Conclusions

There are greater differences in gene expression and molecular processes involved in the progression from normal healthy mucosa to IM than from IM to gastric cancer. While antigen processing is common in both IM-subtypes progressing to GC, more oncogenic processes are observed in the progression of IIM.

Introduction

Gastric carcinogenesis proceeds through a series of precursor lesions in the gastric mucosa named the Correa's cascade, constituted by multi-atrophic gastritis, intestinal metaplasia (IM) and dysplasia conducting to gastric cancer (GC) [1]. In this process, IM is a crucial lesion, due to its high progression rate to GC (3.77/1000 person-years, in the province of Soria, Spain) [2]. IM is a trans-differentiation process of the gastric epithelium to an intestinal type, mostly induced by *H.pylori* infection and expression of the homeobox genes *CDX1* and *CDX2*. It is a protective response against inflammation but it also increases the risk of neoplastic transformation [3].

Intestinal metaplasia has been histologically divided in two types. The complete (CIM, type I, small intestine) is characterized by goblet, absorptive enterocytes and Paneth cells. The incomplete (IIM, types II and III, colonic) is characterized by the presence of goblet and hybrid columnar intermediate cells, absence of enterocytes and rare presence of Paneth cells [2]. Follow-up epidemiologic studies show a higher progression rate to GC of incomplete compared to the complete type of IM [4]. Clinical management guidelines recommend gastroscopy every 3 years when extensive IM is present [5]. However, only a minority of patients with this lesion develop GC [4] and the genes and pathways responsible for this progression are unknown. Therefore, the identification of deregulated genes and molecular processes responsible for this transition is relevant because it could reveal driver genes for tumor progression as well as potential new biomarkers and therapeutic targets.

Gene expression profiling of IM has identified an up-regulation of intestinal differentiation [6,7] and lipid metabolism [6,7,8] genes. However, to our knowledge, there are no published studies comparing samples of IM that progress to gastric cancer along time with those that do not progress. Our hypothesis is that some genes and molecular processes are deregulated in complete and incomplete IM that progress to gastric cancer (CIM-GC and IIM-GC). To test this, we performed expression profiling to compare mRNA from histological subtypes of IM that after a follow-up study had progressed to GC (CIM-GC and IIM-GC) with the mRNA from those IM subtypes that did not progress to GC (CIM-NoGC, IIM-NoGC). Normal gastric mucosa was also compared to IM subtypes that did not progress to GC (IM-NoGC).

Materials and methods

Patients and samples

Samples were obtained from two Spanish follow-up studies of gastric carcinogenesis. One was performed in the province of Soria [2] and the other was a multicenter study including 9

Spanish hospitals [9]. Furthermore, samples from the Gregorio Marañón's hospital, in Madrid, were also included. Patients diagnosed at recruitment with CIM and IIM were subjected to a new gastroscopy after a mean follow-up of 12 ± 3.4 years.

As previously published, formalin-fixed paraffin-embedded (FFPE) samples at recruitment and at the end of follow-up were diagnosed by histology [2,9]. Briefly, 3–4 biopsies from the antrum, incisura or corpus, were stained with hematoxylin–eosin, Alcian blue–periodic acid Schiff (AB-PAS, pH 2.5) and Giemsa. Additionally, for some IIM samples, some sections were stained with high-iron diamine–Alcian blue to detect sulfated mucins [2]. Complete IM was classified by the presence of brush border cells and goblet cells, but without non-goblet and hybrid columnar “intermediate” cells. Incomplete IM, which also contains goblet cells, was diagnosed by the presence of its predominant cell type, the hybrid “intermediate” non-goblet mucous columnar cells. This hybrid mucous cell is easily identified because they show, with the mucino histochemical AB-PAS, a mixed gastric and intestinal phenotype pattern: red for neutral gastric mucins and blue for a combination of intestinal mucins [2]. Two pathologists (MLP and JMSA) reviewed the biopsies of all the included cases. Furthermore, when FFPE blocks were cut for RNA extraction (see below) additional hematoxylin–eosin staining was performed at initial and final cuts to confirm that the tissue in between was from the corresponding IM subtype.

All samples used in this study were obtained from projects approved by the Ethics committee of the Biomedical Research Institute of Bellvitge (CEIC HUB-ICO-IDIBELL), as well as those of the hospitals involved in the projects FIS Exp PI030077 (Hospital de Soria), FIS ExpPI10/01089, PI10/01031 and PI10/01203 and the Gregorio Marañón hospital.

RNA extraction and microarray analysis

Total RNA was extracted from FFPE cuts of gastric biopsies at recruitment. An initial and a final slide for Hematoxylin–Eosin (HE) staining was obtained and evaluated before RNA extraction in order to further confirm the lesion-type and to select those samples that contained at least 50% of IM in all cuts. RNA was obtained from 8 CIM-GC, 6 IIM-GC, 9 CIM--NoGC, 7 IIM-NoGC, and 15 normal gastric mucosae. We selected FFPE blocks of IM (IIM or CIM)-GC patients of similar sex and age as IM (IIM or CIM)-NoCG patients. The characteristics of the patients' samples are shown in [S1 Table](#).

Total RNA was extracted with Recover All Nucleic Acid for FFPE kit (Ambion, USA). RNA quality was evaluated using NanoDrop 2000 (Thermo Scientific, USA) for quantification and purity check, and using the Bioanalyzer 2100 (Agilent Technologies, USA) for RNA integrity analysis. For microarray analysis, 100ng of input RNA from each sample was processed using the SensationPlus FFPE Amplification and 3'IVT Labeling Kit (Affymetrix, USA) and hybridized to the Almac Xcel array (Affymetrix, USA), specifically optimized for use with degraded FFPE samples.

Quality control, statistical and bioinformatics analyses

After quality assessment, data normalization was performed by the Robust Multichip Analysis method ([Affymetrix[®] Expression Console[™] Software](#)) [10] to correct background, apply quantile normalization and summarize all the probes of a transcript in logarithmic base 2. Batch effect was corrected by method Combat [11]. Of the initial 81.804 probe sets, we selected 57.263 above the 30 percentile to eliminate noise produced by probes that do not express enough intensity/signal. A dendrogram representation of the sample hierarchical clustering was constructed using the Euclidean distance and method Complete.

Differentially expressed genes were obtained by moderated t test for microarray data (Limma) [12] and adjusted by False Discovery Rate (FDR) [13]. Three types of comparisons

were performed: IIM-GC vs IIM-NoGC, CIM-GC vs CIM-NoGC and IM-NoGC vs Healthy mucosa. In order to exclude potential tumoral processes, in the comparison of IM with healthy mucosa we did not include IM samples that had progressed to GC. Genes were considered differentially expressed if the t-test p -value was <0.05 and the fold change (FC) ≥ 2 (up-regulated) or ≤ 0.5 (down-regulated). Differences between the analyzed groups in clinical and morphological variables of the study patients (age, sex, *H.pylori* infection status and anatomical location of the lesion in the stomach) were analyzed by means of the t-test for continuous variables and the Fisher exact test for categorical variables. Significant differences ($p < 0.05$) were only obtained for age between IM-NoGC group and healthy mucosa. In agreement with the age and sex matched selection of FFPE blocks from IM-GC and IM-NoGC patients, no age differences were observed between the different groups of intestinal metaplasia (IIM or CIM) patients. There were also no differences between groups for any of the other variables analyzed. Thereafter, p -values for moderated t test were only age adjusted in the comparisons IIM-GC vs Healthy, IIM-NoGC vs Healthy, CIM-GC vs Healthy, CIM-NoGC vs Healthy, IM-NoGC vs Healthy and IM-GC vs Healthy. Analyses were performed with R (v3.1.1) and Bioconductor [14] packages. Data are accessible in GEO (GSE78523).

We constructed Venn diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to identify the Differentially Expressed Genes (DEGs) that were unique or common to the different IIM and CIM analysis groups (IIM-GC vs IIM-NoGC, IM-GC vs Healthy mucosa and IM-NoGC vs Healthy mucosa; and the same for CIM) and to select DEGs specific for IIM-GC vs IIM-NoGC and IIM-GC vs Healthy comparisons, but absent in IIM-NoGC vs Healthy.

We used the GEO2R tool of Gene expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) to analyze the deposited GEO data from a similar study (GSE69146) [15] and to compare with our results.

Gene Set Enrichment Analysis, GSEA

Normalized expression values were loaded to GSEA [16] and the gene set collections c2.all.v5.0 and c3.tft.v5.0 of MSigDB v5.0 [16] were interrogated. c2.all.v5.0 includes gene sets from signaling pathways and genetic or chemical perturbations, while c3.tft.v5.0 contains genes that share a transcription factor binding site in TRANSFAC v7.4. The default parameters were applied with the exception of IIM-GC vs IIM-NoGC where permutations were performed on *Gene set* following methodology recommendations [17]. A gene set was considered significant if the p -value was <0.05 and q -value was <0.25 for CIM-GC vs CIM-NoGC and IM-NoGC vs Healthy. However, for IIM-GC vs IIM-NoGC a gene set was considered significant if the nominal p -value was <0.01 and the q -value was <0.05 to avoid the higher rate of false positives typically encountered when gene sets are used for permutation [18].

We categorized gene sets to molecular processes according to their functions (e.g. COLLER_MYC_TARGETS_UP (oncogenes)). Furthermore, we looked at the intersections between: A) gene sets with enrichment scores (ES) higher than cutoffs (IIM-GC: $ES > 0.5$, IM-NoGC: $ES > 0.65$), B) extreme values of “Rank at max” parameter and C) leading edge genes that are also DEGs [17].

An expression dataset of IM and healthy gastric mucosa [19] deposited in Gene Expression Omnibus (GSE47797) was downloaded and analyzed by GSEA following the previous methodology.

Ingenuity Pathway Analysis (IPA)

The default Core Analysis of IPA [20] was used to identify biological processes, functions, canonical pathways and molecular networks from differentially expressed genes in the three

specified comparisons. This software compares DEGs from an experiment against genes collected in the Ingenuity® Knowledge base. In this way, canonical pathways and molecular networks are identified, providing insight of how DEGs interact [20]. Results with p-value < 0.05 were considered significant.

Validation by RT-qPCR of differentially expressed genes

Validation of 19 DEGs from the microarray was performed by RT-qPCR using the dynamic expression arrays Biomark HD 96x96 (Fluidigm, USA) and UPL probes (Roche, Switzerland). The RT-qPCR protocols (PN 68000116 B3, PN-100-6472 A1, PN 100-5876 B1 and PN 68000130 E1) of Fluidigm were followed. We selected some top up and down-regulated significant genes of interest from Tables 1–3 (S2 Table) and performed duplicated qPCR in CIM-GC (N = 24), CIM-NoGC (N = 33), IIM-GC (N = 11), IIM-NoGC (N = 8) and Healthy (N = 16)

Table 1. Differentially expressed genes with oncogenic functions in the IIM progressing to GC (IIM-GC).

Gene Symbol ^a	Function	Fold Change	Nominal p-value	Associated with gastric carcinogenesis ^b
<i>IK</i>	Antigen Processing	0.384	0.002	New
<i>HLA-A</i>		2.059	0.029	Yes [23]
<i>HLA-DQB1</i>		2.17	0.001	Yes [24]
<i>CD24</i>		2.249	0.008	Yes [25]
<i>HLA-DRB4</i>		2.25	0.024	Yes [24]
<i>HLA-DQA1</i>		3.182	0.024	Yes [24]
<i>HLA-DRB1/3/5</i>		3.368	0.048	Yes [24]
<i>HLA-C</i>		4.417	0.011	Yes [26]
<i>CXCL14</i>	Inflammation	2.007	0.028	Yes, GC prognosis [27]
<i>IL1R2</i>		2.31	0.019	New
<i>BPIFB1</i>		2.866	0.009	New
<i>C1R</i>	Complement system	2.048	0.005	New
<i>C1QBP</i>		2.076	0.002	New
<i>C3</i>		2.193	0.020	New
<i>PPIA</i>	Chaperones	0.393	0.027	Yes [28]
<i>CCT6A</i>		2.053	0.002	New
<i>CANX</i>		2.117	0.027	New
<i>HSP90AA1</i>		2.412	0.0001	Yes, GC prognosis [29]
<i>HSP90AB1</i>		2.479	0.003	Yes, GC prognosis [29]
<i>RBBP7</i>	Tumor suppression	0.449	0.001	New
<i>EIF5B</i>		2.059	0.004	New
<i>CAV1</i>		2.144	0.0005	YES, GC metastasis [30]
<i>EIF3D</i>		2.211	0.003	New
<i>CDKL3</i>	Cell cycle regulator	0.403	0.0001	New
<i>CP</i>	Metabolism of iron	0.412	0.040	Yes, decreased in GC [31]
<i>MYOF</i>	Angiogenesis	2.035	0.006	New
<i>ANP32B</i>	Anti-apoptotic factor	2.155	0.006	New
<i>NHP2</i>	Telomerase	2.187	0.002	Yes, GC prognosis [32]
<i>GNL3</i>	Stem cell proliferation	2.218	0.0003	New
<i>RAN</i>	Oncogene	2.41	0.004	Yes, SNPs associated [33]

^a Genes are ordered increasingly according to Fold Changes inside the functional groups.

^b It refers to genes previously associated with gastric carcinogenesis by genetic association, expression, proteomic or functional studies. IIM, Incomplete Intestinal Metaplasia. GC, Gastric cancer

<https://doi.org/10.1371/journal.pone.0176043.t001>

samples. Some of these samples were the same as those used for the microarray analysis (S1 Table); the remaining samples were from an independent series (S3 Table).

ACTB, *GAPDH*, *G6PD*, *RPL29* and *B2M* genes were considered as potential reference genes due to their lack of significance in the microarray analysis. The selected reference gene was *ACTB*, which was the one that exhibited the lowest significance by ANOVA (p-value = 0.395), when comparing Ct values of IM-GC vs IM-Not GC vs Healthy controls, using the software GraphPad Prism v5.01 [21]. The expression levels of the target genes and their statistical significance were calculated by use of the Bootstratio software [22] after introduction of $2^{-\Delta Ct}$ values for cases and controls, where $\Delta Ct = Ct \text{ mean}(\text{target gene}) - Ct \text{ mean}(\text{ACTB})$. In all cases a p-value < 0.05 was considered significant.

Results

Hierarchical clustering and principal component analysis of the analysed samples based on the gene expression profiles obtained after microarray analysis showed a clear separation between healthy controls and all the IM samples; however, little differences between IM samples that had progressed to GC with respect to those that did not progress were observed (Fig 1). A similar representation of the data was observed in the heatmaps of the compared groups (Fig 2). Differentially expressed genes were observed from the three comparisons, but the differences between IM (either IIM or CIM)-GC and (IIM or CIM)-NoGC were not significant after applying the FDR test.

IIM-GC vs IIM-NoGC: 106 genes were identified as differentially expressed, of which 75.4% were over-expressed and 24.6% under-expressed in the IIM-GC (S4 Table).

CIM-GC vs CIM-NoGC: 19 genes were differentially expressed, being the majority (N = 11) under-expressed in the CIM-GC (S5 Table).

To select tumour progression genes specific for the IMs-GC, we compared the DEG of IM (IIM and CIM) that progressed to GC versus those that did not progress with the DEG in the IMs that did not progress versus healthy mucosa and those in the IMs that progressed to GC versus healthy mucosa (Fig 3). Among the genes specific of IIM or CIM progression to GC, there were more deregulated transcripts with oncogenic functions in the IIM-GC (Table 1) than in the CIM-GC (Table 2).

IIM vs CIM: Comparison of IIM-GC vs CIM-GC revealed 17 DEGs (p-val < 0.05), of which 15 were upregulated and 2 were downregulated (S6 Table), but as in the comparisons of

Table 2. Differentially expressed genes with oncogenic functions in the CIM progressing to GC (CIM-GC).

Gene Symbol ^a	Function	Fold Change	Nominal p-value	Association with gastric carcinogenesis ^b
<i>IGHG1/IGHM/IGHV4-31</i>	Antigen Processing	0.493	0.014	YES
<i>HLA-DRB4</i>		2.049	0.022	YES [24]
<i>HLA-DRB1/3/5</i>		3.475	0.021	YES [24]
<i>ANAPC5</i>	Mitotic factor	0.461	0.031	New
<i>IL1R2</i>	Inflammation	0.475	0.017	New
<i>PRSS1</i>	Protein degradation	0.495	0.033	New
<i>GP2</i>	Innate immunity response	2.063	0.018	New
<i>HOXA13</i>	Intestinal differentiation	2.076	0.028	YES [34]
<i>IGFBP5</i>	Cellular proliferation	2.127	0.004	New
<i>OLFM4</i>	Antiapoptotic factor	3.332	0.0002	YES [35]

^a Genes are ordered increasingly according to Fold Changes inside the functional groups.

^b It refers to genes previously associated with gastric carcinogenesis by genetic association, expression, proteomic or functional studies. CIM, Complete Intestinal Metaplasia. GC, Gastric cancer

Table 3. Differentially expressed genes representative of molecular processes in the IM not progressing to GC (IM-NoGC).

Gene Symbol ^a	Function ^b	Fold Change	p-value	Adj p-value	
<i>CDX2</i>	Intestinal differentiation	3.133	4.050E-11	2.343E-08	
<i>CDX1</i>		4.147	3.523E-09	1.187E-06	
<i>HOXB13</i>	Intestinal differentiation (New DEGs)	2.099	2.508E-04	1.187E-02	
<i>HOXB6</i>		2.884	2.313E-07	4.386E-05	
<i>HOXA13</i>		3.294	7.134E-05	4.632E-03	
<i>MUC3A</i>	Mucins	13.277	1.677E-16	6.001E-13	
<i>MUC12</i>		16.382	4.336E-09	1.395E-06	
<i>MUC17</i>		15.579	1.550E-15	3.773E-12	
<i>MUC2</i>		24.748	5.216E-25	1.857E-20	
<i>APOA1</i>	Lipid Metabolism	7.146	4.060E-07	6.902E-05	
<i>APOA4</i>		7.331	2.059E-08	5.385E-06	
<i>MTTP</i>		23.334	1.176E-14	1.871E-11	
<i>APOB</i>		35.139	5.209E-15	9.479E-12	
<i>FABP1</i>		67.381	7.752E-25	1.857E-20	
<i>CYP3A4</i>		Xenobiotic Metabolism	7.304	4.204E-08	1.003E-05
<i>UGT2A3</i>	2.741		9.798E-06	9.690E-04	
<i>GSTA1</i>	3.285		2.794E-04	1.287E-02	
<i>GAST</i>	Hypochlorhidria	6.249	4.603E-10	2.075E-07	
<i>SST</i>		0.183	1.815E-08	4.789E-06	
<i>SLC46A3</i>	Transporters (New process)	5.553	1.458E-10	7.258E-08	
<i>SLC7A9</i>		6.007	7.421E-11	4.047E-08	
<i>SLC17A4</i>		7.240	3.905E-12	3.105E-09	
<i>SLC5A1</i>		8.238	2.024E-09	7.492E-07	
<i>SLC13A2</i>		8.412	2.326E-14	3.330E-11	
<i>SLC6A19</i>		9.009	1.581E-15	3.773E-12	
<i>SLC26A3</i>		16.877	3.539E-11	2.136E-08	
<i>TRIM15</i>		NFkB activation(New DEGs)	2.256	4.063E-07	6.902E-05
<i>TRIM31</i>			2.372	8.101E-06	8.298E-04
<i>TRIM36</i>			2.548	3.530E-07	6.125E-05
<i>TRIM40</i>	2.677		1.287E-04	7.379E-03	
<i>SNORD116s</i>	Expression regulators (New process)	0.443	1.378E-04	7.675E-03	
<i>TMEM25</i>	TMEM function (New process)	2.768	1.617E-04	8.622E-03	
<i>TMEM139</i>		3.386	2.229E-10	1.064E-07	
<i>TMEM45B</i>		2.709	1.380E-09	5.450E-07	
<i>DMBT1</i>	Tumor suppressor	71.951	9.727E-25	1.857E-20	

^a Genes are ordered increasingly according to Fold Changes inside the functional groups.

^b New means that novel DEGs or molecular processes were found. IIM, Incomplete Intestinal Metaplasia. GC, Gastric cancer

<https://doi.org/10.1371/journal.pone.0176043.t003>

IIM-GC vs IIM-NoGC and CIM-GC vs CIM-NoGC, none of them was significant after FDR adjustment for multiple comparisons (FDR). Nevertheless, it is to note the increased expression in IIM-GC of immunoglobulin and inflammatory gene products, indicating an increased active immunoinflammatory response in IIM-GC. There was also an increased expression in IIM-GC of some key molecules in the gastric function such as pepsinogen, H⁺/K⁺ channel generator of the H⁺ ion for HCl synthesis and gastric intrinsic factor.

Comparison of IIM-NoGC vs CIM-NoGC revealed 16 DEGs (p-val<0.05), of which four DEGs were upregulated and 11 were downregulated (S7 Table); again, none of them was

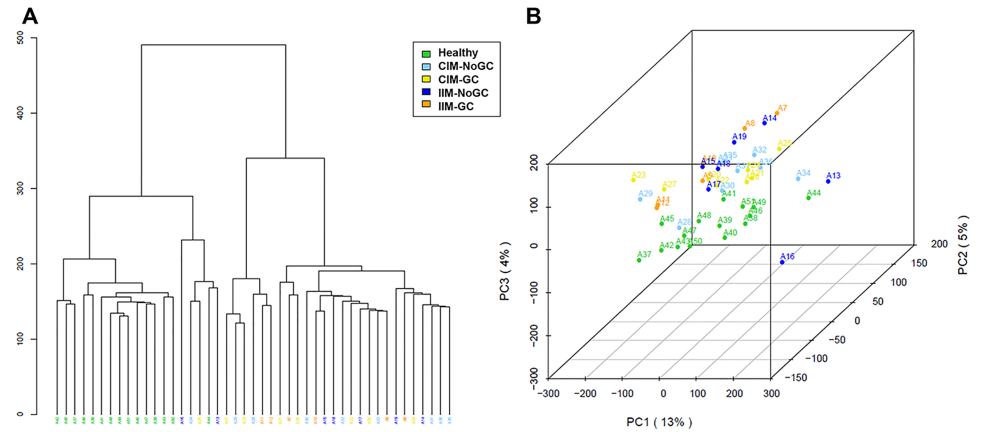


Fig 1. Dendrogram and principal component analysis of the analysed samples. A) Dendrogram showing the hierarchical clustering of the analysed samples. Clustering was based on the overall gene expression values of the studied groups. B) Principal Components Analysis.

<https://doi.org/10.1371/journal.pone.0176043.g001>

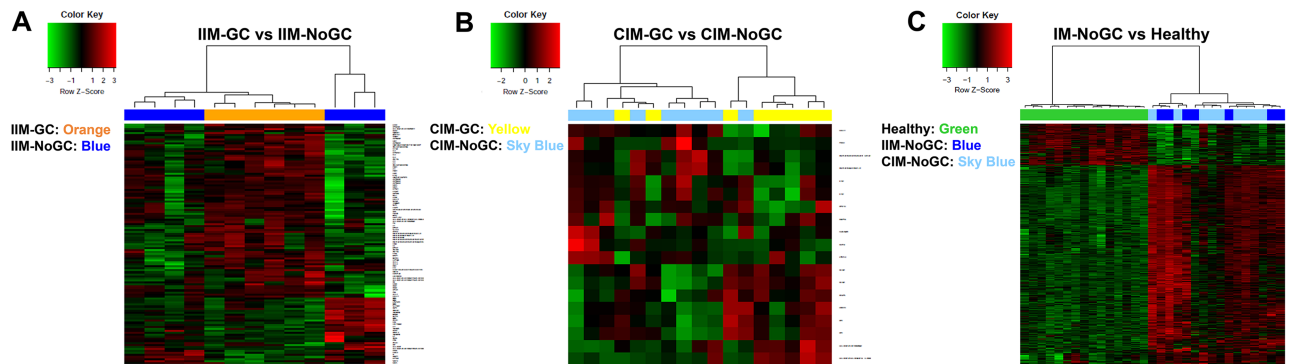


Fig 2. Heat maps of the analyzed groups. A) IIM-GC vs IIM-NoGC, B) CIM-GC vs CIM-NoGC, C) IIM-NoGC vs Healthy.

<https://doi.org/10.1371/journal.pone.0176043.g002>

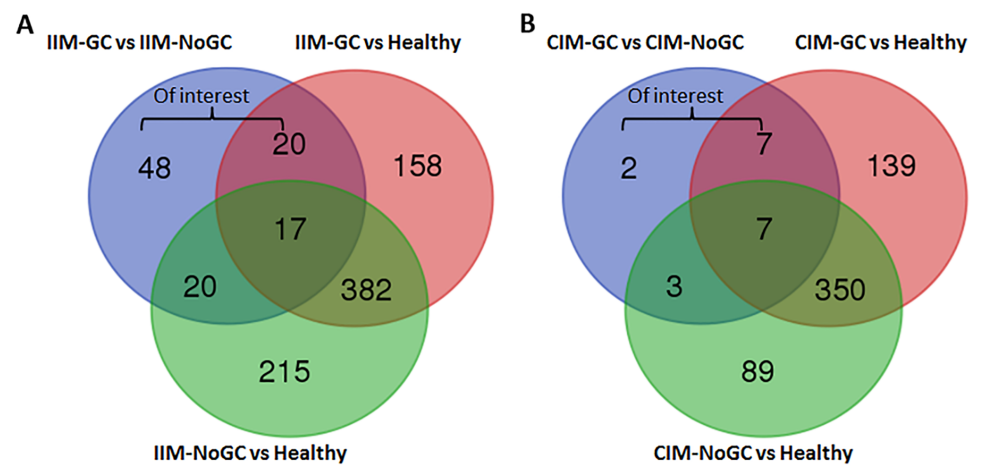


Fig 3. Venn diagrams of the differentially expressed genes in three different comparisons. A, Incomplete intestinal metaplasia. B, Complete intestinal metaplasia. IIM-GC or CIM-GC, incomplete or complete intestinal metaplasia progressing to gastric cancer. IIM-NoGC or CIM-NoGC, incomplete or complete intestinal metaplasia not progressing to gastric cancer. Healthy, healthy gastric mucosa.

<https://doi.org/10.1371/journal.pone.0176043.g003>

significant after FDR adjustment. Five of these DEGs were the same as in IIM-GC vs CIM-GC but four of them (*CXCL17*, *PGC* and two corresponding to *IGH* genes) were upregulated in IIM-GC and downregulated in IIM-NoGC, while *EBF1* was downregulated in both comparisons.

IIM-NoGC vs Healthy: We identified 482 DEGs (394 upregulated and 88 downregulated), all of which significant after FDR test. With respect to previous comparisons it is relevant to note the higher number of both DEGs and expression levels ($FC > 4$ $N = 92$) and that they are all significant after FDR (S8 Table). DEGs contributing to molecular processes previously reported in IM were confirmed and new genes contributing to these processes as well as to new processes were also identified (Table 3).

Gene Set Enrichment Analysis (GSEA)

From the c2all.v5 gene set collection we obtained 144 gene sets significantly over-expressed in the IIM-GC. From the c3tft.v5 collection we obtained 19 overexpressed gene sets, all composed by E2F translation initiation factors. The S9 Table indicates the number of significant gene sets per molecular process after categorization according to their function. To select the most relevant molecular processes represented by these gene sets, cut-offs to *Rank at max* and *ES* parameters were applied and up-regulated leading edge genes were identified (S11 and S12 Tables). Comparison of these results with those of the gene sets with enrichment scores above 0.5 indicated that activation of cell cycle and cell proliferation, oncogenes, tumor suppressors and insulin regulated genes were important processes in the progression of IIM to GC (Fig 4A).

For the CIM-GC group we only obtained two significant gene sets (GAZDA_DIAMOND_-BLACKFAN_ANEMIA_ERYTHROID_UP, REACTOME NETRIN1_SIGNALING) without any apparent relationship with gastric carcinogenesis.

When both IM subtypes that do not progress to GC were grouped and compared with healthy mucosa, 120 gene sets grouped in different molecular processes were over-expressed in IM-NoGC (S10 Table). There were 12 over-expressed gene sets from the c3.tft.v5.0 catalogue composed by transcription factors HNF1/4 and GATA1/6, which cooperate to induce intestinal genes [3].

When we compared the relevant molecular processes in IM-NoGC, following identical methodology as indicated for IIM-GC (S13 and S14 Tables), the most important processes were found to be Warburg effect, lipid metabolism, intestinal differentiation, aberrant glycosylation, antigen processing, inflammation and apoptosis (Fig 4B).

Ingenuity Pathway Analysis (IPA)

The IPA analysis showed that, unlike to the GSEA, there were several up-regulated pathways common to the IIM-GC and CIM-GC such as antigen presentation, signaling by TNFRSF4,

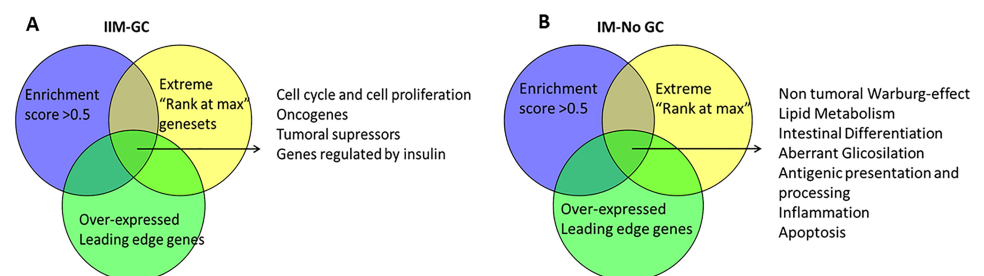


Fig 4. Venn’s diagram to select the most relevant molecular processes after GSEA analyses. A, Incomplete intestinal metaplasia progressing to gastric cancer (IIM-GC). **B,** Both types of intestinal metaplasia not progressing to gastric cancer (IM-NoGC).

<https://doi.org/10.1371/journal.pone.0176043.g004>

Table 4. Over-expressed canonical pathways and other information from Ingenuity Pathway Analysis.

Group	Canonical pathway	p-value	Upstream Regulators	Diseases	Cellular and molecular functions
IIM-GC	Antigen presentation	3.15E-10	HOXC11	-Immunological	-Cell proliferation
IIM-GC	TNFRSF4 (OX40) signaling	3.27E-09	RAD21	-Connective Tissue	-Cell death and survival
IIM-GC	Thyroid autoimmune disease	1.66E-08	EBI3	-Inflammatory	-Cellular movement
IIM-GC	Development of B cells	1.51E-07	NLRC5	-Muscular and Skeletal	-Cellular development
IIM-GC	Maturation of dendritic cells	1.26E-06	SMC3	-Development	-Cell-cell interaction and signaling
IIM-GC	Phagosome maturation	1.59E-06			
IIM-GC	Communication between innate and adaptive cells	1.62E-06			
CIM-GC	Communication between innate and adaptive cells	3.14E-05	LGALS3	-Endocrine	-Cell-cell interaction and signaling
CIM-GC	Maturation of dendritic cells	3.22E-04	LHCGR	-Gastrointestinal	-Cellular movement
CIM-GC	Antigen presentation	3.98E-04	MYBL2	-Immunological	-Cell death and survival
CIM-GC	TNFRSF4 (OX40) signaling	7.49E-04	mir-296	-Inflammatory	-Cell cycle
CIM-GC			CHI3L1	-Development	- Cellular assembly and organization
IM-NoGC	Activation of FXR/RXR α	1.42E-08	HNF4A	-Endocrine	-Cellular movement
IM-NoGC	Activation of PXR/RXR α	1.34E-07	HNRNPA2B1	-Metabolism	-Death and cell survival
IM-NoGC	Inhibition of RXR α by LPS/IL1	2.24E-05	HNF1A	-Gastrointestinal	-Lipid metabolism
IM-NoGC	Development of B cells	7.76E-05	CTNBB1	-Hepatic	-Molecular transport
IM-NoGC	Degradation of sucrose	4.30E-05	SREBF1	-Cardiovascular	-Biochemistry of small molecules, drugs and amino acids
IM-NoGC	Metabolism of thyroid hormone	0.0004			
IM-NoGC	Hematopoiesis of stem cells	0.0005			
IM-NoGC	Activation of LXR/RXR	0.0006			
IM-NoGC	Degradation of melatonin (New)	0.0013			
IM-NoGC	Glycolysis I	0.0039			
IM-NoGC	Xenobiotic Metabolism Signaling	0.0052			

<https://doi.org/10.1371/journal.pone.0176043.t004>

communication between innate and adaptive cells, maturation of dendritic cells and the possible role of other homeobox genes such as *HOXC11* in the development of the intestinal differentiation. Due to the small number of differentially expressed genes in CIM-GC (Table 2), pathways over-expressed in this group were composed by a small number of genes and of lower statistical significance, when compared with the IIM-GC group (Table 4). Regarding IM-NoGC, IPA confirmed the GSEA results for metabolism of lipids and xenobiotics and a previously observed deregulation of the thyroid hormone metabolism. A new process of melatonin degradation was also identified (Table 4).

The molecular network with highest score in the IIM-GC was found to be composed of effectors of the immune response such as HLA class I (HLA-A, HLA-C) and II (DQB1, DRB1) molecules, the cytokine receptor IL1R2, immunoglobulins, and oncogenic molecules such as the chaperones CANX and HSP90AA1 and the tumor suppressor CAV1 (S1 Fig). There was not a high score molecular network in CIM-GC. The molecular network with the highest score in the IM-NoGC includes the processes of lipid metabolism, molecular transport and biochemistry of small molecules (S2 Fig).

Validation of DEGS from microarray results by RT-qPCR

RT-qPCR was used to validate selected DEGs obtained in the IIM-GC (N = 7), CIM-GC (N = 4) and IM-NoGC (N = 11) groups (S2 Table). Only 51% of reactions passed the quality control defined by the Fluidigm software, probably reflecting RNA degradation. Of the 19 DEGs selected from the microarray, 9 were also differentially expressed in the qPCR, but only

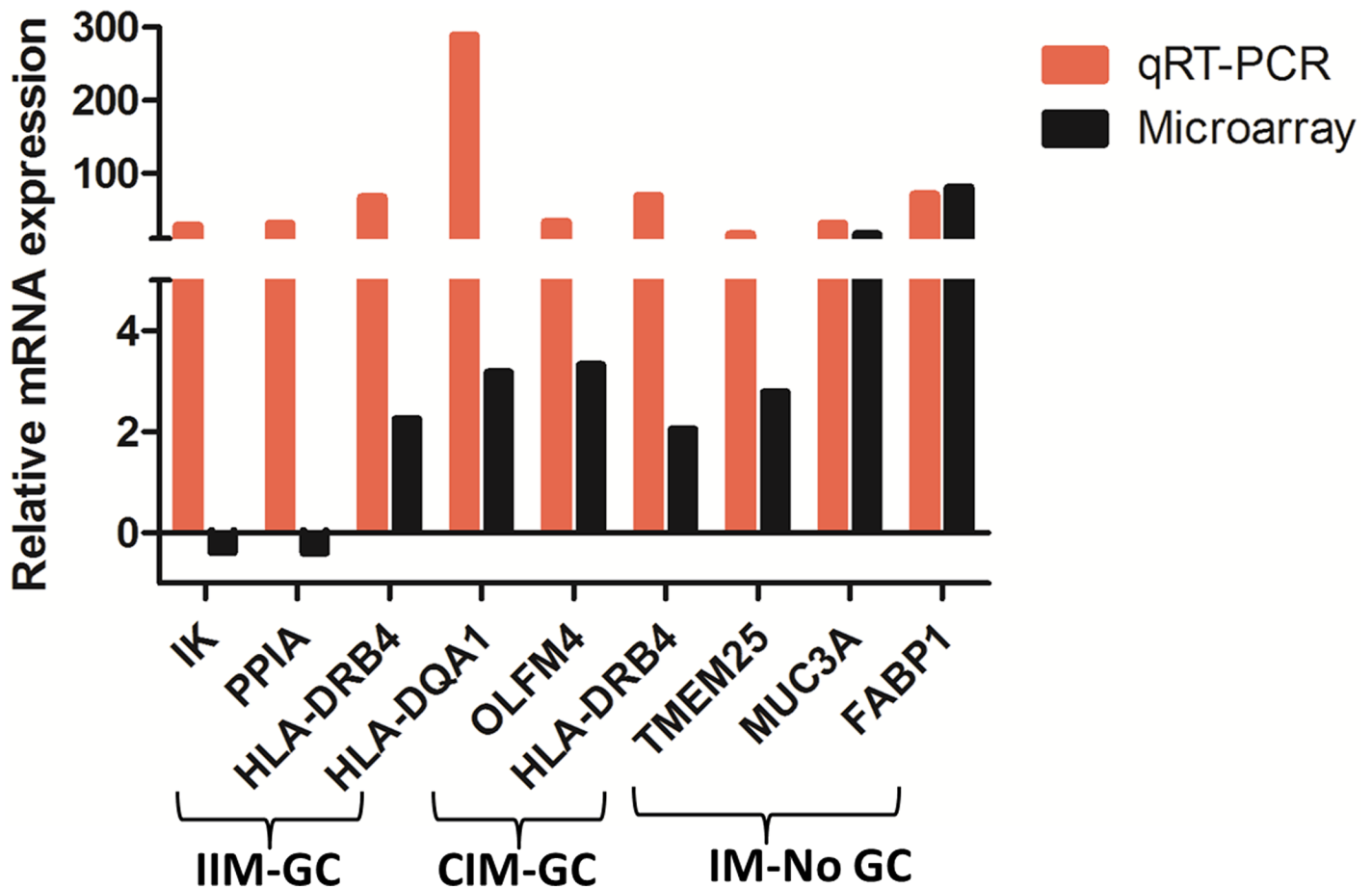


Fig 5. Validation by RT-qPCR of some differentially expressed genes obtained in the microarray analysis. Comparison of the expression level (fold change) by microarray analysis and by RT-qPCR of significant differentially expressed genes in the IIM-CG versus IIM-NoGC, CIM-GC versus CIM-NoGC and IM-NoGC versus healthy gastric mucosa.

<https://doi.org/10.1371/journal.pone.0176043.g005>

in 7 of them (36.84% of total DEGs) the difference in expression had the same direction (Fig 5). One of the validated genes both in CIM-GC and IMM-GC is *HLA-DRB4*, which is in agreement with the fact that antigen presentation is a molecular process common to the progression of both IM subtypes in the GSEA and IPA analyses. Also, the highest up-regulated gene in CIM-GC, *OLFM4*, was validated.

Discussion

As far as we know, this is the first study on differential gene expression between IM subtypes that have progressed to gastric cancer and those who have not progressed, after a several years follow-up period. Our results from gene expression and pathway analysis indicate that IM-NoGC is very different from healthy gastric mucosa but the differences between IM subtypes that progress to GC and those that do not progress are smaller and of lower magnitude. Despite this general finding, we have identified DEGs and molecular processes involved in the progression from IM subtypes to gastric cancer. Furthermore, we have confirmed previously reported molecular processes that distinguish IM from healthy gastric mucosa and have identified new transcripts contributing to these processes.

Although expression differences were not found significant after multiple corrections when IIM-GC was compared to IIM-NoGC, some of the DEGs have been previously associated with gastric carcinogenesis. There are several up-regulated transcripts reflecting the inflammatory process induced by *H. pylori* infection, such as *HLA-A*, *HLA-C* [26], *HLA-DRB1/3/5*, *HLA-DQA1/DQB1* [24, 36]. It is interesting to note that *HLA-DRB* is up-regulated in both IIM-GC and CIM-GC, that *HLA-DRB4* overexpression was confirmed by RT-qPCR and that antigen presentation also is a common pathway in both IM-subtypes progressing to GC, altogether being in agreement with the immune response as an important process in the progression of precursor lesions to GC [37]. There are several up-regulated inflammatory transcripts (*CXCL14*, *IL1R2*, *BPIFB1*) and also from the complement system (*C1R*, *C1QB*, *C3*) which increase inflammation and phagocytosis of bacteria, respectively. Regarding this, other complement components such as *C1S*, *C1QR1* and *CD55* have been found over-expressed in GC [38]. Several chaperones responsible for proper oncoproteins folding were found overexpressed, such as *CCT6A*, *CANX*, *HSP90AA1* and *HSP90AB1*. Interestingly, over-expression of *HSP90AA1* and *HSP90AB1* has been associated with poor prognosis in GC [29]. The regulators of the tumor suppressor *RB1*, *EIF3D* and *EIF5B* were also found over-expressed, as well as *NHP2*, a telomerase whose high expression in GC is correlated with poor clinical prognosis [32]. A member of RAS oncogenes called RAN, with genetic polymorphisms associated with of GC risk [33], was also found overexpressed. Novel over-expressed genes not previously associated with gastric carcinogenesis are *MYOF*, an angiogenesis regulator [39], and *GNL3*, that interacts with TP53 and may be involved in carcinogenesis because of its role in cancer and stem cells proliferation [40].

Apart from the above mentioned overexpression of *HLA-DRB*, in the CIM-GC group there were fewer significant transcripts than in IIM-GC. Furthermore, it is to note the increased expression in IIM-GC of immunoglobulin (*IGHG*, *IGHM*) and inflammatory (*CXCL17*) gene products, indicating an increased active immunoinflammatory response in IIM-GC when compared to CIM-GC. All these results are in agreement with the increased progression risk to GC of IIM compared with CIM [4]. It is likely that oncogenic pathways only become activated after CIM progresses to IIM along the Correa's cascade. Among the over-expressed genes with oncogenic functions in CIM-GC there are *OLFM4* (olfactomedin 4), whose overexpression was validated by RT-qPCR, and *IGFBP5*, an anti-apoptotic and a cell proliferation factor, respectively. *OLFM4* is increased in early stages of gastric carcinogenesis and a prognostic marker for advanced GC [35]. *OLFM4* has also been found to promote tumor growth in pancreatic cancer [41], altogether suggesting that it could be an early factor of gastric tumor progression from CIM.

Confirmation of our main gene expression results through immunohistochemistry would be the best proof of their validity. Unfortunately, we had no more FFPE sections for further immunohistochemical staining. Therefore, to further validate our results by comparison with other available datasets from similar studies, we have searched PubMed and the Gene Expression Omnibus (GEO) databases for microarray studies exploring the same or similar progression phenotype. We found a recent study (GEO accession: GSE69144) that used microarray analysis to determine gene expression changes in the progression along the gastric precancerous cascade of gastric lesions, in *H. pylori* infected subjects followed for a period of 6 years [15]. However, comparison of this study with ours is difficult because of clear differences in study design and microarray platform used. While this study was designed for the identification of changes in gene expression after lesion progression or regression, our study was designed for the identification of biomarkers of progression to GC at baseline samples. Furthermore, in our study we analyzed a high-density transcriptome based microarray containing 97,000 transcripts, while Garay et al. used a Cancer Panel Array restricted to 502 genes. Therefore, only two of the differentially expressed genes of interest in IM progressing to GC in our study (Tables 1 and 2) were also present in the GSE69144 dataset: *CAV1* and *IGFBP5*. Comparison

with the GEO2R tool of those baseline samples of the Garay study that had progressed with those that did not change revealed that *IGFBP5* is found similarly overexpressed as in the comparison of CIM-GC vs CIM-NoGC in our study. This gene has also been reported to be up-regulated in gastric cancer [42]. *CAVI* was not significant in Garay's study.

To characterize the transcriptional profile of IM in comparison to healthy gastric mucosa, we analyzed samples of both histological subtypes that do not progress to GC because these are more common and also to exclude tumorigenic processes from the analysis. Unlike previous comparisons, DEGs in IM were significant after multiple comparisons and exhibited greater expression differences. There are many DEGs previously found up-regulated in IM, such as *CDX2*, *KRT20*, *MUC13*, *OLFM4*, *REG4* [43], *FABP1*, *MEP1B*, *SI*, *SLC6A19* [44], *CDX1*, *MTTP*, *CEACAM6* [6], *APOA1*, *APOA4*, *APOB*, *CDH17*, *CLDN3*, *HNF4A*, *VIL1* [7]. Their functions show that intestinal differentiation, metabolism of lipids and xenobiotics, and hypochlorhidria, characterize the IM. New DEGs identified in the IM-NoGC group reveal the role of other homeobox genes (*HOXA13/B6/C*) in the intestinal differentiation, transporters of the solute carrier family (SLC), NFKB activators (TRIM), small nucleolar RNAs (SNORD116) and TMEM proteins of unknown function. The highest over-expressed gene is *DMBT1* (S6 Table), a GC tumor suppressor and immunohistochemical marker of IM [45]. It may be speculated that its high expression could be an early defense mechanism against a potential oncogenic deregulation along time.

The results from GSEA also showed a difference between IIM-GC and CIM-GC; in the first group there were many over-expressed gene sets with oncogenic functions, while none was identified in the CIM-GC. The most relevant molecular processes in IIM-GC are antigen processing, inflammation, activation of cell cycle and cell proliferation, oncogenes and tumor suppressors. Other published molecular processes support these differences between the IM subtypes, indicating that IIM is a more advanced state than CIM along the Correa's cascade [1]. Thus, when compared to CIM, IIM has decreased expression of *CDX2* [46], less activation of Sonic Hedgehog pathway [47], higher microsatellite instability [48], increased telomerase activity [49] and increased intracellular localization of *H.pylori* [50].

Regarding the IM-NoGC group, among the most relevant processes we found the Warburg effect, which consists that in tumoral processes glycolysis is increased with respect to oxidative phosphorylation [51] [52]. In IM-NoGC there is an over-expression of glucose transporters *SLC2A5*, *SLC5A9*, the glycolytic enzyme hexokinase and glucagon, a promoter of gluconeogenesis. The final product of glycolysis, lactate, activates the hypoxia inducible factors *HIF1A* and *HIF2A*, which induce glycolytic genes [52], and in this regard there are two over-expressed *HIF1A* gene sets. Warburg effect is universal in normal cells under active proliferation to protect them from genomic damage [52], and this could be its role in IM-NoGC. As already indicated by the DEGs analysis, other relevant processes in IM-NoGC were found to be lipid metabolism and the O-glycosylation of mucins. During active cell proliferation, biological membranes with high lipid contents, are synthesized for new cells [53]. Besides, since lipids are used as a gluconeogenic energy source, increased lipid metabolism could be at least in part a consequence of Warburg effect. Mucins are major components of gastric mucus, with high O-glycosylation contributing to mucosa protection from bacterial infections. *H.pylori* modulates glycosylation to create a micro-environment favorable to infection. In the IM, aberrant glycosylation of *MUC5AC* and *MUC6* increase their hydro-solubility [54], maybe allowing penetration of *H.pylori* in the gastric mucosa and hence increasing inflammation [50].

The consistency of our results is also indicated by the general agreement obtained when we compared by GSEA the expression profile of IM-NoGC with the raw data of a microarray expression dataset of IM and healthy gastric mucosa performed by Hanada *et al* [19] and

deposited in GEO. As in our study, major biological processes over-expressed in that study were non tumoral Warburg effect, lipid metabolism, intestinal differentiation, transcription factors HNF1A/4A, inflammation, aberrant protein glycosylation, apoptosis, xenobiotic metabolism and response to genomic damage.

The results from IPA confirmed the findings of previous works and from GSEA about an up-regulation of antigen presentation in both IM subtypes that progress to GC, as well as metabolism of lipids, xenobiotics and dysregulation of thyroid hormone [55] in IM not progressing to GC. New results from this analysis indicate that other transcriptional factors different from CDX1/CDX2, such as HOXC11, could induce intestinal differentiation. HOXC11 is an upstream regulator and other members of this family such as HOXA13, HOXB6/7/13 are over-expressed in IM. Other new up-regulated molecular process in IM not progressing to GC obtained from IPA is melatonin degradation.

The novelties of this study are that we analyzed both metaplasia histological subtypes that progress or not to GC, with a high extension of the lesion (above 75% in most samples) and performed a functional enrichment analysis based on GSEA and IPA. However, the small sample size, particularly in the CIM-GC and IIM-GC groups, is the main limitation of our study since it reduces its power to obtain significant results after correction for multiple comparisons. Small sample size, together with the fact that most (75%) of the samples were reported to be negative for *H.pylori* infection, was also the main reason for not performing a stratified analysis by *H.pylori* infection. It is nevertheless to note that obtaining appropriate samples for expression analysis of IM subtypes whose progression or not to GC is known from follow-up studies is not an easy task. The results of this study are therefore a starting point for the design of new and larger studies aimed at their validation.

To conclude, our results are the first to indicate that the transcriptional profile of IM subtypes that progress to GC shows small differences in the gene expression levels in comparison with those IM subtypes that do not progress. Antigen presentation and inflammation are up-regulated processes common to both IM histological subtypes that progress to GC but the IIM subtype shows a higher number of up-regulated oncogenic DEGs and molecular processes than CIM, which is in agreement with its higher risk of progression to GC.

In agreement with previous reports, the transcriptional profile of IM-NoGC with respect to healthy mucosa evidences a drastic difference in gene expression. Apart from already reported processes such as intestinal differentiation, metabolism of lipids and xenobiotics, new molecular processes observed in this study are non tumoral Warburg effect and melatonin degradation. Newly identified transcripts include TRIM, TMEM, homeobox genes from HOX family, transporter genes and the small nucleolar RNAs, SNORDs116.

Supporting information

S1 Fig. Molecular network with highest score in the IIM-GC comparison, after Ingenuity Pathway Analysis.

(TIF)

S2 Fig. Molecular network with highest score in the IM-NoGC comparison, after Ingenuity Pathway Analysis.

(TIF)

S1 Table. Main characteristics of the samples analyzed in the expression microarray.

(DOC)

S2 Table. Significant genes from the expression microarray selected for validation by qPCR.

(DOC)

S3 Table. Independent series of samples used in the validation by qPCR.
(DOC)

S4 Table. Differentially expressed genes in IIM-GC.
(DOC)

S5 Table. Differentially expressed genes in CIM-GC.
(DOC)

S6 Table. Differentially expressed genes between IIM-GC and CIM-GC.
(DOC)

S7 Table. Differentially expressed genes between IIM-NoGC and CIM-NoGC.
(DOC)

S8 Table. Differentially expressed genes in the IM-NoGC.
(DOC)

S9 Table. GSEA analysis in IIM-GC from c2all.v5 and c3tft.v5 catalogs.
(DOC)

S10 Table. GSEA analysis in IM-NoGC from c2all.v5 and c3tft.v5 catalogs.
(DOC)

S11 Table. Over-expressed gene sets in IIM-GC with extreme values of *Rank at max* parameter.
(DOC)

S12 Table. Gene sets composed by at least 3 leading edge genes up-regulated in the IIM-GC.
(DOC)

S13 Table. Over-expressed gene sets in IM-NoGC with extreme values of *Rank at max* parameter.
(DOC)

S14 Table. Gene sets composed by at least 3 leading edge genes overexpressed in the IM-NoGC.
(DOC)

Acknowledgments

We acknowledge Miguel Lopéz Heredia, Virginia Nunes and Clara Vilches for their help in the design and performance of qPCR experiments, as well as Alvaro Aytés for his comments and critical review of the manuscript.

Author Contributions

Conceptualization: OC JMS CAG NS.

Data curation: OC LN.

Formal analysis: OC LN NS.

Funding acquisition: CAG JMS JG NS.

Investigation: OC EP LN NG JMS JG CAG NS.

Methodology: OC JMS EP LN NS.

Resources: OS NG EP LN JMS MLP VP CL VA MC MG JG.

Supervision: CAG NS.

Validation: OC NS JMS MLP.

Visualization: OC NS.

Writing – original draft: OC NS.

Writing – review & editing: OC JMS MLP EP LN NG VP CL VA MC MG JG CAG NS.

References

- Correa P, Piazzuelo MB. The gastric precancerous cascade. *J Dig Dis*. 2012; 13(1):2–9. <https://doi.org/10.1111/j.1751-2980.2011.00550.x> PMID: 22188910
- González CA, Pardo ML, Ruiz Liso JM, Alonso P, Bonet C, Garcia RM, et al. Gastric cancer occurrence in preneoplastic lesions: A long-term follow-up in a high-risk area in Spain. *Int J Cancer*. 2010; 127(11):2654–60. <https://doi.org/10.1002/ijc.25273> PMID: 20178099
- Mesquita P, Raquel A, Nuno L, Reis CA, Silva LFS, Serpa J, et al. Metaplasia—a transdifferentiation process that facilitates cancer development: the model of gastric intestinal metaplasia. *Crit Rev Oncog*. 2006; 12(1–2):3–26. PMID: 17078204
- González CA, Sanz-Anquela JM, Gisbert JP, Correa P. Utility of subtyping intestinal metaplasia as marker of gastric cancer risk. A review of evidence. *Int J Cancer*. 2013; 133(5):1023–32. <https://doi.org/10.1002/ijc.28003> PMID: 23280711
- Dinis-Ribeiro M, Areia M, De Vries AC, Marcos-Pinto R, Monteiro-Soares M, O'Connor A, et al. Management of precancerous conditions and lesions in the stomach (MAPS): Guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHS), European Society of Pathology (ESP), and the Sociedade Portuguesa. *Virchows Arch*. 2012; 460(1):19–46. <https://doi.org/10.1007/s00428-011-1177-8> PMID: 22190006
- Boussioutas A, Li H, Liu J, Waring P, Lade S, Holloway AJ, et al. Distinctive Patterns of Gene Expression in Premalignant Gastric Mucosa and Gastric Cancer. *Cancer Res*. 2006; 63(10):2569–77.
- Chen X. Variation in Gene Expression Patterns in Human Gastric Cancers. *Mol Biol Cell*. 2003; 14(8):3208–15. <https://doi.org/10.1091/mbc.E02-12-0833> PMID: 12925757
- Gomes LI, Esteves GH, Carvalho AF, Cristo EB, Hirata R, Martins WK, et al. Expression Profile of Malignant and Nonmalignant Lesions of Esophagus and Stomach: Differential Activity of Functional Modules Related to Inflammation and Lipid Metabolism. *Cancer Res*. 2005; 65(16):7127–36. <https://doi.org/10.1158/0008-5472.CAN-05-1035> PMID: 16103062
- González CA, Sanz Anquela JM, Companioni O, Bonet C, Berdasco M, López C, et al. Incomplete type of intestinal metaplasia has the highest risk to progress to gastric cancer: results of the Spanish follow-up multicenter study. *J Gastroenterol Hepatol*. 2015; 31(5):953–8.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4(2):249–64. <https://doi.org/10.1093/biostatistics/4.2.249> PMID: 12925520
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8(1):118–27. <https://doi.org/10.1093/biostatistics/kxj037> PMID: 16632515
- Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. 2005; 21(9):2067–75. <https://doi.org/10.1093/bioinformatics/bti270> PMID: 15657102
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995; 57(1):289–300.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5(10):R80. <https://doi.org/10.1186/gb-2004-5-10-r80> PMID: 15461798
- Garay J, Piazzuelo MB, Majumdar S, Li L, Trillo-Tinoco J, Del Valle L, et al. The homing receptor CD44 is involved in the progression of precancerous gastric lesions in patients infected with *Helicobacter pylori* and in development of mucous metaplasia in mice. *Cancer Lett*. 2016; 371(1):90–8. <https://doi.org/10.1016/j.canlet.2015.10.037> PMID: 26639196

16. Subramanian A, Subramanian A, Tamayo P, Tamayo P, Mootha VK, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517
17. Gene Set Enrichment Analysis GSEA User Guide. The Broad Institute. <http://software.broadinstitute.org/gsea/index.jsp>
18. Sonachalam M, Shen J, Huang H, Wu X. Systems Biology Approach to Identify Gene Network Signatures for Colorectal Cancer. *Front Genet*. 2012; 3:80. <https://doi.org/10.3389/fgene.2012.00080> PMID: 22629282
19. Hanada K, Uchida T, Tsukamoto Y, Watada M, Yamaguchi N, Yamamoto K, et al. Helicobacter pylori Infection Introduces DNA Double-Strand Breaks in Host Cells. *Infect Immun*. 2014; 82(10):4182–9. <https://doi.org/10.1128/IAI.02368-14> PMID: 25069978
20. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014; 30(4):523–30. <https://doi.org/10.1093/bioinformatics/btt703> PMID: 24336805
21. Prism version 5.01 for Windows, GraphPad Software [Internet]. La Jolla California USA; www.graphpad.com
22. Clèries R, Galvez J, Espino M, Ribes J, Nunes V, de Heredia M. BootstRatio: A web-based statistical analysis of fold-change in qPCR and RT-qPCR data using resampling methods. *Comput Biol Med*. 2012; 42(4):438–45. <https://doi.org/10.1016/j.combiomed.2011.12.012> PMID: 22270228
23. Mimura K, Shiraishi K, Mueller A, Izawa S, Kua L-F, So J, et al. The MAPK Pathway Is a Predominant Regulator of HLA-A Expression in Esophageal and Gastric Cancer. *J Immunol*. 2013; 191(12):6261–72. <https://doi.org/10.4049/jimmunol.1301597> PMID: 24244023
24. Magnusson PKE, Enroth H, Eriksson I, Held M, Nyrén O, Engstrand L, et al. Gastric cancer and human leukocyte antigen: distinct DQ and DR alleles are associated with development of gastric cancer and infection by Helicobacter pylori. *Cancer Res*. 2001; 61(6):2684–9. PMID: 11289148
25. Wu J-X, Zhao Y-Y, Wu X, An H. Clinicopathological and prognostic significance of CD24 overexpression in patients with gastric cancer: a meta-analysis. *PLoS One*. 2014; 9(12):e114746. <https://doi.org/10.1371/journal.pone.0114746> PMID: 25503963
26. Ishigami S, Arigami T, Uenosono Y, Matsumoto M, Okumura H, Uchikado Y, et al. Cancerous HLA class I expression and regulatory T cell infiltration in gastric cancer. *Cancer Immunol Immunother*. 2012; 61(10):1663–9. <https://doi.org/10.1007/s00262-012-1225-5> PMID: 22374482
27. Xue X. Abnormal hypermethylation of promoter region downregulates chemokine CXC ligand 14 expression in gastric cancer. *Int J Oncol*. 2013; 43(5):1487–94. <https://doi.org/10.3892/ijo.2013.2078> PMID: 23982764
28. Bai Z, Ye Y, Liang B, Xu F, Zhang H, Zhang Y, et al. Proteomics-based identification of a group of apoptosis-related proteins and biomarkers in gastric cancer. *Int J Oncol*. 2011; 38(2):375–83. <https://doi.org/10.3892/ijo.2010.873> PMID: 21165559
29. Wang J, Cui S, Zhang X, Wu Y, Tang H. High expression of heat shock protein 90 is associated with tumor aggressiveness and poor prognosis in patients with advanced gastric cancer. *PLoS One*. 2013; 8(4):e62876. <https://doi.org/10.1371/journal.pone.0062876> PMID: 23638161
30. Nam KH, Lee BL, Park JH, Kim J, Han N, Lee HE, et al. Caveolin 1 expression correlates with poor prognosis and focal adhesion kinase expression in gastric cancer. *Pathobiology*. 2013; 80(2):87–94. <https://doi.org/10.1159/000341685> PMID: 23038627
31. Aquino PF, Fischer JSG, Neves-Ferreira AGC, Perales J, Domont GB, Araujo GDT, et al. Are gastric cancer resection margin proteomic profiles more similar to those from controls or tumors? *J Proteome Res*. 2012; 11(12):5836–42. <https://doi.org/10.1021/pr300612x> PMID: 23145836
32. Kim MS, Kim SS, Yoo NJ, Lee S. Expressional analysis of NOLA1, NOLA2, NOLA3 and DKC1, the core proteins in H/ACA riboproteins, in gastric and colorectal cancers. *Pathology*. 2012 Oct; 44(6):576–7. <https://doi.org/10.1097/PAT.0b013e3283583bf6> PMID: 22935986
33. Xie Y, Wang Y, Zhao Y, Guo Z. Single-nucleotide polymorphisms of microRNA processing machinery genes are associated with risk for gastric cancer. *Onco Targets Ther*. 2015; 8:567–71. <https://doi.org/10.2147/OTT.S79150> PMID: 25784816
34. Han Y, Tu W-W, Wen Y-G, Li D-P, Qiu G-Q, Tang H-M, et al. Identification and validation that up-expression of HOXA13 is a novel independent prognostic marker of a worse outcome in gastric cancer based on immunohistochemistry. *Med Oncol*. 2013; 30(2):564. <https://doi.org/10.1007/s12032-013-0564-1> PMID: 23592225
35. Jang BG, Lee BL, Kim WH. Olfactomedin-related proteins 4 (OLFM4) expression is involved in early gastric carcinogenesis and of prognostic significance in advanced gastric cancer. *Virchows Arch*. 2015; 467(3):285–94. <https://doi.org/10.1007/s00428-015-1793-9> PMID: 26070873

36. Wee A, Teh M, Kang JY. Association of *Helicobacter pylori* with HLA-DR antigen expression in gastritis. *J Clin Pathol*. 1992; 45(1):30–3. PMID: [1740510](#)
37. Lee K, Hwang H, Nam KT. Immune response and the tumor microenvironment: How they communicate to regulate gastric cancer. *Gut Liver*. 2014; 8(2):131–9. <https://doi.org/10.5009/gnl.2014.8.2.131> PMID: [24672653](#)
38. Zang S, Guo R, Xing R, Zhang L, Li W, Zhao M, et al. Identification of Differentially-expressed Genes in Intestinal Gastric Cancer by Microarray Analysis. *Genomics Proteomics Bioinformatics*. 2014; 12(6):276–83. <https://doi.org/10.1016/j.gpb.2014.09.004> PMID: [25500430](#)
39. Bernatchez PN, Acevedo L, Fernandez-Hernando C, Murata T, Chalouni C, Kim J, et al. Myoferlin regulates vascular endothelial growth factor receptor-2 stability and function. *J Biol Chem*. 2007; 282(42):30745–53. <https://doi.org/10.1074/jbc.M704798200> PMID: [17702744](#)
40. Tsai RYL, McKay RDG. A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. *Genes Dev*. 2002; 16(23):2991–3003. <https://doi.org/10.1101/gad.55671> PMID: [12464630](#)
41. Kobayashi D, Koshida S, Moriai R, Tsuji N, Watanabe N. Olfactomedin 4 promotes S-phase transition in proliferation of pancreatic cancer cells. *Cancer Sci*. 2007; 98(3):334–40. <https://doi.org/10.1111/j.1349-7006.2007.00397.x> PMID: [17270022](#)
42. Chen Y, Guo Y, Ge X, Itoh H, Watanabe A, Fujiwara T, et al. Elevated expression and potential roles of human Sp5, a member of Sp transcription factor family, in human cancers. *Biochem Biophys Res Commun*. 2006; 340(3):758–66. <https://doi.org/10.1016/j.bbrc.2005.12.068> PMID: [16380080](#)
43. Lee HJ, Nam KT, Park HS, Kim MA, LaFleur BJ, Aburatani H, et al. Gene Expression Profiling of Metaplastic Lineages Identifies CDH17 as a Prognostic Marker in Early Stage Gastric Cancer. *Gastroenterology*. 2010; 139(1):358–66.
44. Kim KR, Oh SY, Park UC, Wang JH, Lee JD, Kweon HJ, et al. Gene expression profiling using oligonucleotide microarray in atrophic gastritis and intestinal metaplasia. *Korean J Gastroenterol*. 2007; 49(4):209–24. PMID: [17464166](#)
45. Sousa JF, Ham AJL, Whitwell C, Nam KT, Lee HJ, Yang HK, et al. Proteomic profiling of paraffin-embedded samples identifies metaplasia-specific and early-stage gastric cancer biomarkers. *Am J Pathol*. 2012; 181(5):1560–72. <https://doi.org/10.1016/j.ajpath.2012.07.027> PMID: [22944598](#)
46. Liu Q, Teh M, Ito K, Shah N, Ito Y, Yeoh KG. CDX2 expression is progressively decreased in human gastric intestinal metaplasia, dysplasia and cancer. *Mod Pathol*. 2007; 20(12):1286–97. <https://doi.org/10.1038/modpathol.3800968> PMID: [17906616](#)
47. Shiotani A, Iishi H, Uedo N, Ishiguro S, Tatsuta M, Nakae Y, et al. Evidence that loss of sonic hedgehog is an indicator of *Helicobacter pylori*-induced atrophic gastritis progressing to gastric cancer. *Am J Gastroenterol*. 2005; 100(3):581–7. <https://doi.org/10.1111/j.1572-0241.2005.41001.x> PMID: [15743355](#)
48. Hamamoto T, Yokozaki H, Semba S, Yasui W, Yunotani S, Miyazaki K, et al. Altered microsatellites in incomplete-type intestinal metaplasia adjacent to primary gastric cancers. *J Clin Pathol*. 1997; 50(10):841–6. PMID: [9462267](#)
49. Kameshima H, Yagihashi A, Yajima T, Kobayashi D, Denno R, Hirata K, et al. *Helicobacter pylori* infection: Augmentation of telomerase activity in cancer and noncancerous tissues. *World J Surg*. 2000; 24(10):1243–9. PMID: [11071470](#)
50. Necchi V, Candusso ME, Tava F, Luinetti O, Ventura U, Fiocca R, et al. Intracellular, Intercellular, and Stromal Invasion of Gastric Mucosa, Preneoplastic Lesions, and Cancer by *Helicobacter pylori*. *Gastroenterology*. 2007; 132(3):1009–23. <https://doi.org/10.1053/j.gastro.2007.01.049> PMID: [17383424](#)
51. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144(5):646–74. <https://doi.org/10.1016/j.cell.2011.02.013> PMID: [21376230](#)
52. Zhang S, Yang C, Yang Z, Zhang D, Ma X, Mills G, et al. Homeostasis of redox status derived from glucose metabolic pathway could be the key to understanding the Warburg effect. *Am J Cancer Res*. 2015; 5(3):928–44. PMID: [26045978](#)
53. Baenke F, Peck B, Miess H, Schulze A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Dis Model Mech*. 2013; 6(6):1353–63. <https://doi.org/10.1242/dmm.011338> PMID: [24203995](#)
54. Joncquel Chevalier Curt M, Lecointe K, Mihalache A, Rossez Y, Gosset P, Leonard R, et al. Alteration or adaptation, the two roads for human gastric mucin glycosylation infected by *Helicobacter pylori*. *Glycobiology*. 2015; 25(6):617–31. <https://doi.org/10.1093/glycob/cwv004> PMID: [25595949](#)
55. Liu R, Li Z, Bai S, Zhang H, Tang M, Lei Y, et al. Mechanism of cancer cell adaptation to metabolic stress: proteomics identification of a novel thyroid hormone-mediated gastric carcinogenic signaling pathway. *Mol Cell Proteomics*. 2009; 8(1):70–85. <https://doi.org/10.1074/mcp.M800195-MCP200> PMID: [18723843](#)