Supplementary Tables and Figures

Supplementary 1 (S1): Flow chart depicting the participants within the Singapore Gastric Cancer Epidemiology and Molecular Genetics Program (GCEP). 56 who passed away due to causes unrelated to GC. Participants who did not complete year 5 surveillance endoscopy were matched against the National Registry of Diseases Office for missed diagnoses of GC, whereby 1 case of GC was diagnosed 7 years after index endoscopy.

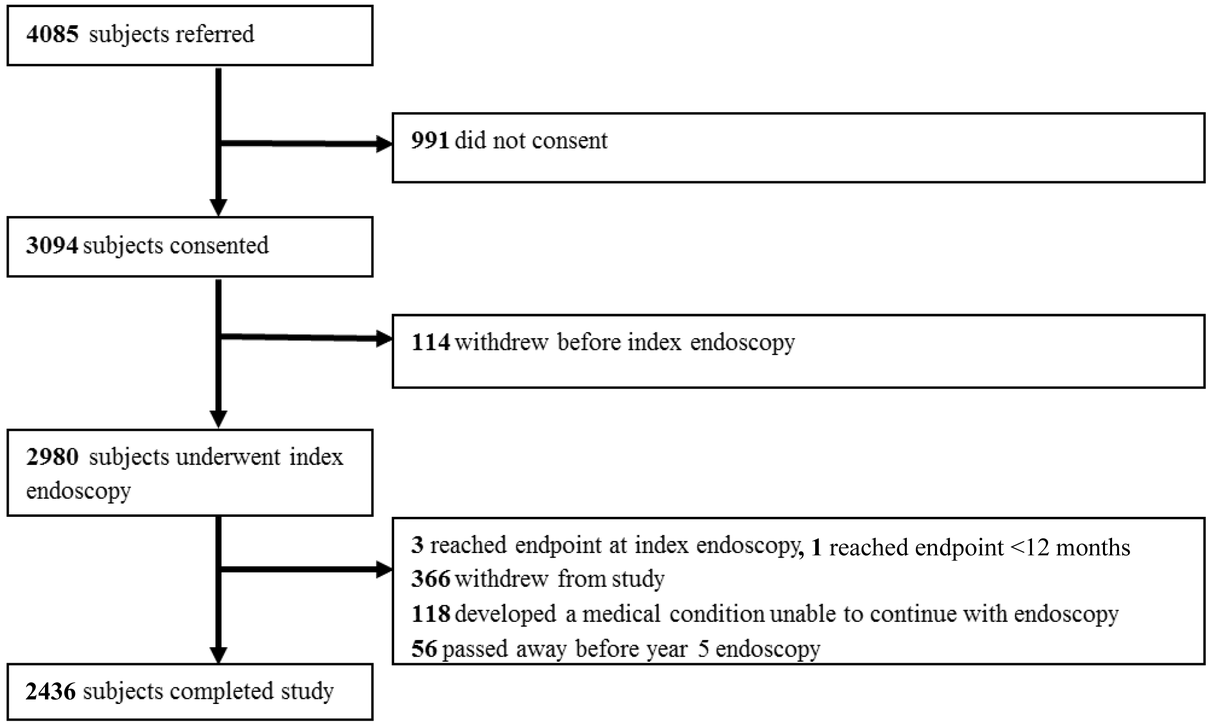
Supplementary 2 (S2): Biopsy protocol according to the updated Sydney System. 5 biopsy sites: One biopsy each from the A1 (antrum lesser curvature), A2 (antrum greater curvature), IA (incisura), B1 (corpus lesser curvature) and B2 (corpus greater curvature) was taken for histological examination. (Ref, Dixon MF et al. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994.) The gastric biopsies were then scored according to the classification of gastritis using OLGIM staging (table shown at the bottom of the figure), whereby the table is adapted from Capelle L et al Gastrointestin Endosc 71: 1150, 2010

Supplementary 3 (S3): Univariate and Multivariate Cox regression analysis to identify risk factors for EGNs

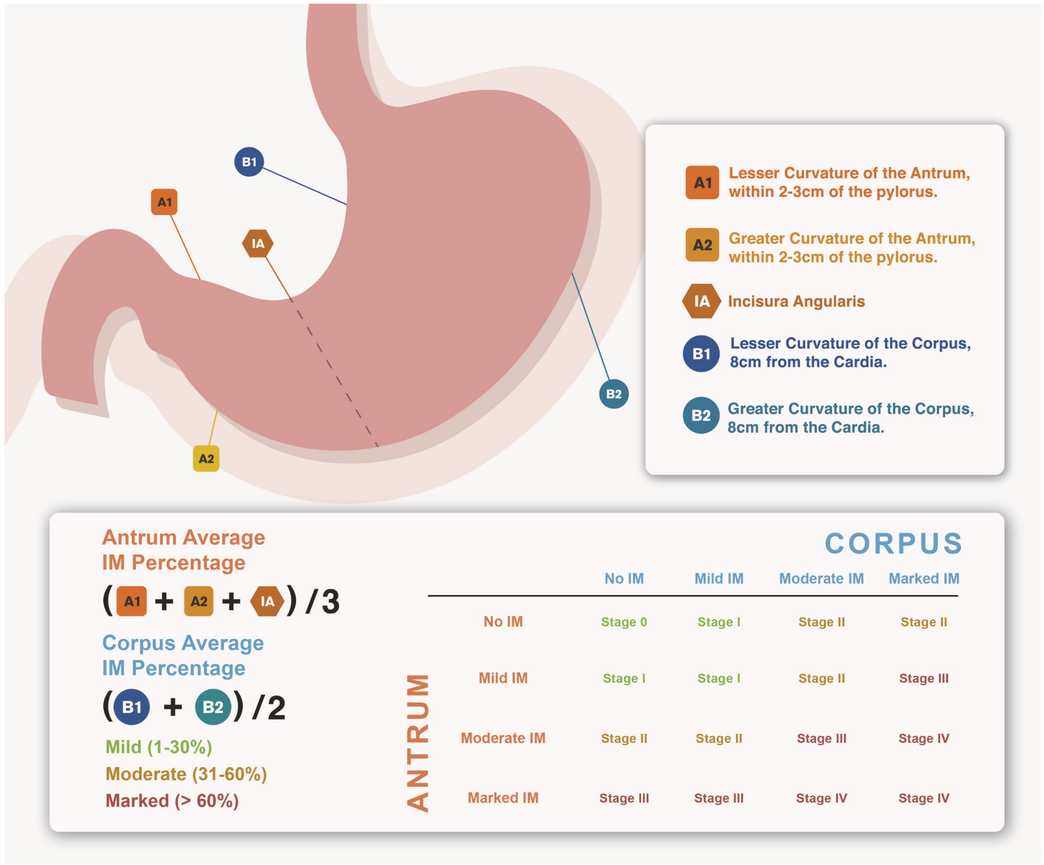
Supplementary 4 (S4): Differential effect of OLGIM II-IV on the risk of EGN progression in subgroups of participants with established risk factors for gastric cancer. The forest plot demonstrates that the presence of baseline OLGIM II-IV lesions remain a significant risk factor for EGN progression, in almost all subgroups. Interaction between the presence of OLGIM II-IV and each subgroup variable was also tested (p-value for interaction), whereby there was no evidence for any of the subgroups participating as a significant effect modifier for OLGIM II-IV on the risk of EGN progression.

Supplementary 5 (S5): (A) Cumulative hazard function for participants with OLGIM II–IV stratified by smoking status defined by pack-years. Participants with OLGIM II–IV (red line) who had >=20 pack-years had statistically significant increased risk of EGN with HR 3.69 (95% CI 1.03–13.2, p=0.045), when compared to non-smokers. Participants with OLGIM II–IV who had <20 pack-years however were not at statistically increased risk with despite HR 2.06 (95%CI 0.41 – 10.3, p=0.38. **(B)** Cumulative hazard function for participants with OLGIM II–IV stratified by either complete subtype (grey) or incomplete subtype (red). Participants with OLGIM II–IV with incomplete subtype (red line) were at increased risk of EGN with HR 5.96 (95%CI 0.77–46.4, p=0.088), when compared to participants with OLGIM II–IV with complete subtype**. (C)** Images showing H&E of complete IM and incomplete IM. **(D)** Effect of incomplete IM subtype and the risk of developing EGN for the subset of participants with moderate or marked IM**.**

Supplementary 6 (S6). Risk Stratification Using Serology Hp-PG Panel to predict IM progression amongst patients with OLGIM I



Supplementary 1 (S1): Flow chart depicting the participants within the Singapore Gastric Cancer Epidemiology and Molecular Genetics Program (GCEP). 56 who passed away due to causes unrelated to GC. Participants who did not complete year 5 surveillance endoscopy were matched against the National Registry of Diseases Office for missed diagnoses of GC, whereby 1 case of GC was diagnosed 7 years after index endoscopy.



Supplementary 2 (S2): Biopsy protocol according to the updated Sydney System. 5 biopsy sites: One biopsy each from the A1 (antrum lesser curvature), A2 (antrum greater curvature), IA (incisura), B1 (corpus lesser curvature) and B2 (corpus greater curvature) was taken for histological examination. (Ref, Dixon MF et al. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994.) The gastric biopsies were then scored according to the classification of gastritis using OLGIM staging (table shown at the bottom of the figure), whereby the table is adapted from Capelle L et al Gastrointestin Endosc 71: 1150, 2010. In brief, each biopsy sample, gastric IM was scored on a four-tiered scale (no IM = 0%, score = 0; mild IM = 1–30%, score = 1; moderate IM = 31–60%, score = 2; Marked IM = > 60%, score = 3).

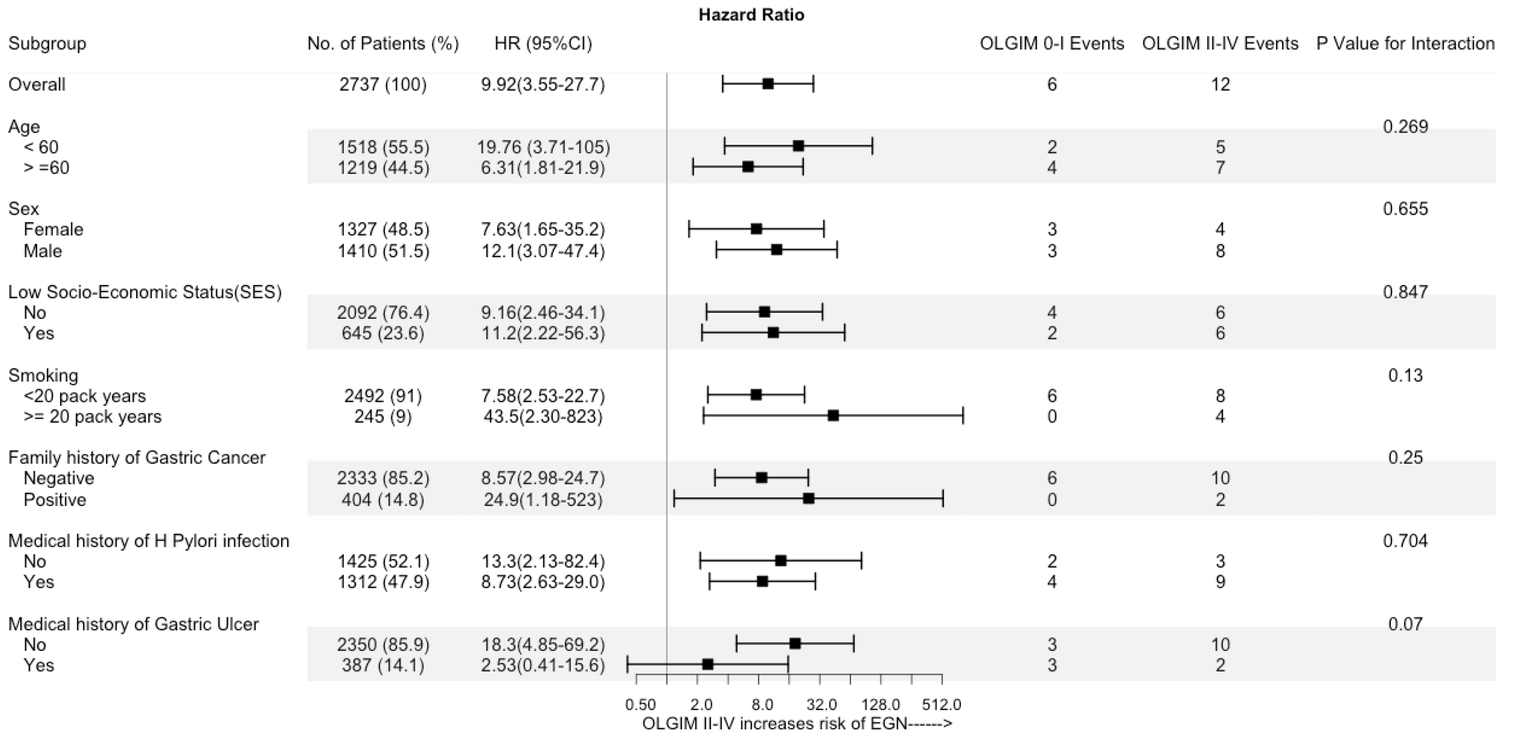
Supplementary 3 (S3): Univariate and Multivariate Cox regression analysis to identify risk factors for EGNs

Risk Factors		Proportion/Mean	No EGN (n=2719)	EGN (n=17)	HR for EGN (univariate)	HR for EGN (multivariate)#
Age (years) Mean (SD)		59.1 (6.7)	59.3 (6.8)	63.6 (8.8)	1.10 (1.03-1.17, p=0.005)	1.08 (1.02-1.16, p=0.016)
BMI Mean (SD)		23.7 (3.6)	23.7 (3.6)	23.5 (3.9)	0.99 (0.86-1.13, p=0.851)	-
Gender	**Female**	1439 (48.3)	1320 (48.5)	7 (41.2)	-	-
	Male	1541 (51.7)	1399 (51.5)	10 (58.8)	1.31 (0.50-3.45, p=0.581)	-
Low SES*	**No**	2245 (75.3)	2082 (76.6)	10 (58.8)	-	-
	Yes	735 (24.7)	637 (23.4)	7 (41.2)	2.71 (1.03-7.13, p=0.044)	1.79 (0.65-4.88, p=0.258)
Alcohol consumption	**Absent**	2936 (98.5)	2262 (83.2)	11 (64.7)	-	-
	Present	44 (1.5)	457 (16.8)	6 (35.3)	2.81 (1.04-7.60, p=0.042)	2.69 (0.85-8.48, p=0.092)
Smoking (pack years) ᵠ	**0**	2318 (77.8)	2139 (78.7)	11 (64.7)	-	-
	0-20	375 (12.6)	339 (12.5)	3 (17.6)	1.93 (0.54-6.92, p=0.313)	1.72 (0.48-6.20, p=0.408)
	>20	287 (9.6)	241 (8.9)	3 (17.6)	2.52 (0.70-9.04, p=0.156)	2.26 (0.63-8.19, p=0.213)
First-degree family history of GC	**Absent**	2558 (85.8)	2317 (85.2)	15 (88.2)	-	-
	Present	422 (14.2)	402 (14.8)	2 (11.8)	0.74 (0.17-3.23, p=0.687)	-
History of H pylori infection	**Absent**	1557 (52.2)	1420 (52.2)	5 (29.4)	-	-
	Present	1423 (47.8)	1299 (47.8)	12 (70.6)	2.75 (0.97-7.83, p=0.057)	2.45 (0.86-6.99, p=0.094)
History of gastric ulcer	**Absent**	2545 (85.4)	2337 (86.0)	12 (70.6)	-	-
	Present	435 (14.6)	382 (14.0)	5 (29.4)	2.71 (0.95-7.69, p=0.061)	2.05 (0.71-5.96, p=0.186)
TFF3 Mean (SD)		6.7 (3.7)	6.8 (4.4)	8.0 (3.8)	1.04 (0.99-1.08, p=0.089)	1.02 (0.95-1.10, p=0.499)
MIF Mean (SD)		31.3 (18.5)	31.3 (18.7)	29.2 (20.0)	1.00 (0.97-1.02, p=0.736)	-
Pepsinogen Index	**Negative**	2521 (93.5)	2491 (93.6)	13 (76.5)	-	-
	Positive	175 (6.5)	169 (6.4)	4 (23.5)	4.52 (1.47-13.86, p=0.008)	4.23 (1.34-13.37, p=0.014)
Atrophic gastritis	**Absent**	2286 (76.7)	2073 (76.2)	9 (52.9)	-	-
	Present	694 (23.3)	646 (23.8)	8 (47.1)	3.16 (1.21-8.20, p=0.018)	2.69 (1.03-7.06, p=0.044)
Intestinal Metaplasia	**Absent**	1659 (55.7)	1557 (57.3)	3 (17.6)	-	
	Present	1321 (44.3)	1162 (42.7)	14 (82.4)	6.87 (1.97-23.99, p=0.003)	5.36 (1.51-19.02, p=0.009)
OLGIM stage	**No IM**	1659 (55.7)	1557 (57.3)	3 (17.6)	-	-
	OLGIM I	906 (30.4)	789 (29.0)	3 (17.6)	2.21 (0.44-10.98, p=0.333)	1.95 (0.39-9.74, p=0.417)
	OLGIM II	252 (8.5)	229 (8.4)	4 (23.5)	9.54 (2.13-42.69, p=0.003)	7.34 (1.60-33.73, p=0.010)
	OLGIM III-IV	163 (5.4)	144 (5.3)	7 (41.2)	29.88 (7.60-117.43, p<0.001)	20.77 (5.04-85.61, p<0.001)

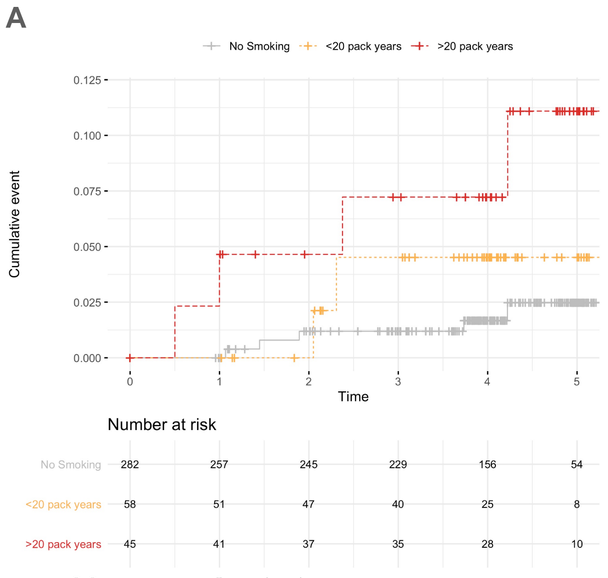
Risk factors with *P* value <0.15 were included in the multivariate regression model to calculate the adjusted hazard ratio and were adjusted for age, low SES and smoking.

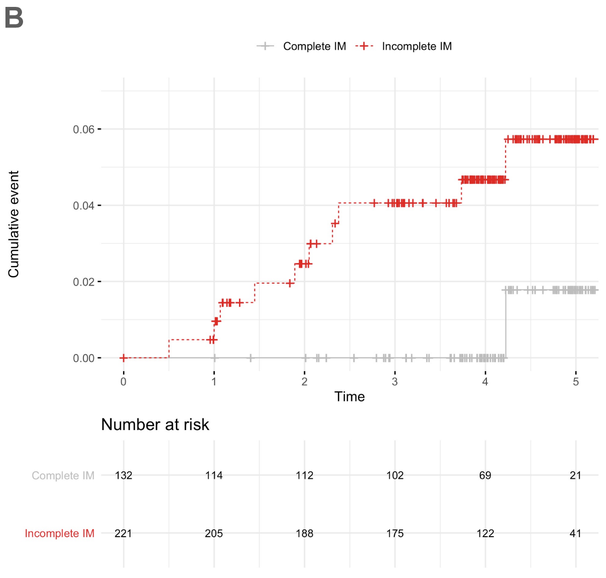
*Low SES (socioeconomic status) is defined as the person has either low education level (primary or below) or the monthly income is below S$1,500.

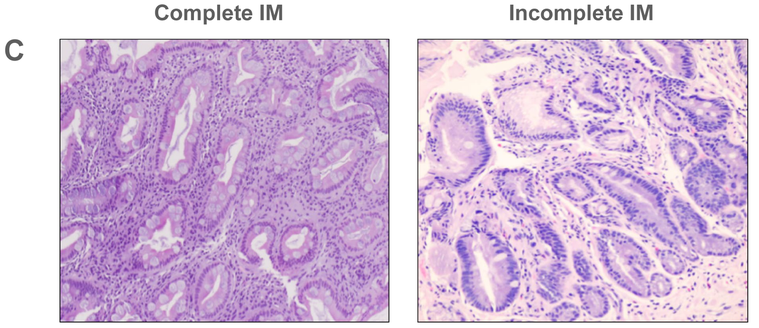
ᵠ The amount of smoking in pack years, which is the number of packs of cigarettes smoked per day multiplied by the number of years of smoking.



Supplementary 4 (S4): Differential effect of OLGIM II-IV on the risk of EGN progression in subgroups of participants with established risk factors for gastric cancer. The forest plot demonstrates that the presence of baseline OLGIM II-IV lesions remain a significant risk factor for EGN progression, in almost all subgroups. Interaction between the presence of OLGIM II-IV and each subgroup variable was also tested (p-value for interaction), whereby there was no evidence for any of the subgroups participating as a significant effect modifier for OLGIM II-IV on the risk of EGN progression.







D

Subtype N=546	No. of Subjects	No. of EGN (%)	No. of non-EGN	Odds Ratio
Complete	302	2 (0.7%)	300	Ref
Incomplete	244	13 (4.9%)	231	8.4 (1.9 - 37.8), p=.005
Total	546	15	531	

(A) Cumulative hazard function for participants with OLGIM II–IV stratified by smoking status defined by pack-years. Participants with OLGIM II–IV (red line) who had >=20 pack-years had statistically significant increased risk of EGN with HR 3.69 (95% CI 1.03–13.2, p=0.045), when compared to non-smokers. Participants with OLGIM II–IV who had <20 pack-years however were not at statistically increased risk with despite HR 2.06 (95%CI 0.41 – 10.3, p=0.38)

(B) Cumulative hazard function for participants with OLGIM II–IV stratified by either complete subtype (grey) or incomplete subtype (red). Participants with OLGIM II–IV with incomplete subtype (red line) were at increased risk of EGN with HR 5.96 (95%CI 0.77–46.4, p=0.088), when compared to participants with OLGIM II–IV with complete subtype.

(C) Images showing H&E of complete IM and incomplete IM.

(D) Effect of incomplete IM subtype and the risk of developing EGN for the subset of participants with moderate or marked IM.

Supplementary 6 (S6). Risk Stratification Using Serology Hp-PG Panel to predict IM progression amongst patients with OLGIM I

Risk Category	**Pepsinogen Index**	**Serum *H. pylori***	**(a) Number of patients n=774**	**Number of patients with progression to OLGIM II-IV or EGN (% of a) n=107**	**HR (95% CI)**
Average Risk (Group A)	Negative	Positive	117	6 (5.1%)	Ref
Moderate Risk (Group B)	Negative	Positive	586	85 (14.5%)	3.78 (1.64 - 8.71), p<.01
High Risk (Group C)	Positive	-	71	16 (22.5%)	5.81 (2.26 - 14.9), p<.01