Genetic variation and gastric cancer risk: a field synopsis and meta-analysis

Simone Mocellin, Daunia Verdi, Karen A. Pooley, Donato Nitti

SUPPLEMENTARY INFORMATION

Materials and Methods

Both intra-gene and inter-gene variants were investigated: as regards the latter group, the nearest gene was considered the relevant gene exclusively for classification purposes (the functional effects of many tested polymorphisms are unknown). The gene symbols used are those recommended by the Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC, http://www.genenames.org).

The following data were extracted from eligible studies: authors' names; region/country where the study was conducted; year of publication; numbers of cases and controls; prevalent ethnicity (>80%); allelic frequency in both cases and controls (if no raw data were available, summary data were collected [i.e., odds ratios and confidence intervals]); MAF and Hardy-Weinberg equilibrium (HWE) in controls; study design (population-based versus hospital-based), genotyping method, HP status, histological subtype (intestinal versus diffuse) and primary cancer site (cardia versus non-cardia).

Whenever possible, the Hardy-Weinberg equilibrium (HWE) in controls was calculated for each study, as per HuGENet recommendations [1].

We also performed evaluations of heterogeneity, subgroup analyses, sensitivity analyses and examination for bias. Between-study heterogeneity (true variance of effect size across studies) was formally investigated by means of Cochran Q-test and I-square statistic (which indicates the percentage of the variability in effect estimates due to true heterogeneity rather than within-study sampling error).

The extent to which the combined risk estimate might be affected by individual studies was assessed by consecutively omitting every study from the meta-analysis (leave-one-out sensitivity

analysis); this approach would also capture the effect of the oldest or first positive study. Publication and selection biases in meta-analysis are more likely to affect small studies, which also tend to be of lower methodological quality: this may lead to "small study effect", where the smaller studies in a meta-analysis show larger effects [2]. A funnel plot was used to detect this effect and asymmetry was formally investigated with the Egger linear regression approach modified by Harbord [3]. The excess of significant findings (potentially indicating the so-called "chasing bias") was evaluated by the test proposed by Ioannidis [4].

Statistical power was estimated based on the ability to detect an OR of 1.15 (or its reciprocal 0.87), with an alpha level equal to the observed P-value.

Results and Discussion

In this synopsis, interesting results came from subgroup meta-analysis. First, we confirmed the difference between the results obtained in populations of different ancestry. In particular, the minor allele frequency (MAF) of studied polymorphisms was often remarkably different between Asian and Caucasian controls, the mean Asian-to-Caucasian MAF ratio being 2.1 (range: 1.1-7.2, see **Supplementary Figure 4**). Then, among 49 significant analyses, 46 (93.9%) were specific for Asians or Caucasians (**Supplementary Figure 5**), suggesting that genetic variation linked to disease risk might be race specific or the LD structure in the different populations may aid in the genetic fine-mapping of these regions in uncovering the most likely causal variants. These observations support the hypothesis that the molecular pathway to gastric cancer susceptibility is not necessarily the same across different ethnicities and underscore the need for race specific risk prediction tools.

As far as high quality biomarkers, only three out of 11 variants were investigated in studies including both Asian and Caucasian populations (PSCA rs2294008, CASP8 rs3834129 and TNF rs1799724), the remaining being tested exclusively in populations of Asian ancestry. This is likely the consequence of the higher number of Asian studies (555/824, 67%) coupled with the fact that all three GWAS were performed in people of Asian ancestry, which in turn probably reflects the special interest in gastric cancer in Asian countries where the disease incidence is higher than in Western countries. These findings call for more investigation on polymorphisms never tested in non-Asian ethnicities.

Focusing on tumor histology (intestinal versus diffuse), subgroup analysis showed that among 16 significant contrasts, 14 (87.5%) regarded exclusively one histotype (**Supplementary Figure 6**). Interestingly, the highest number of significant associations belonged to the diffuse subtype (n=12, 75.0%), and three of the 11 associations with high level of summary evidence regarded this histotype (MUC1 rs2070803, MTX1 rs2075570, PKLR rs3762272), whose incidence is increasing especially in Western countries [5, 6].

Furthermore, considering the site of primary tumor (cardia versus non-cardia), available data showed that, among 14 significant contrasts, 11 (78.6%) were specific to one anatomical location (**Supplementary Figure 7**). In particular, half these associations regarded the non-cardia site, whose incidence is declining especially in Western countries (as opposite to the cardia carcinoma [5, 6]). Also among the 11 high quality variants three were site specific (non-cardia: PSCA rs2294008, PRKAA1 rs13361707; cardia: PLCE1 rs2274223).

Considering the different epidemiology and aggressiveness of different subsets of the disease [5, 6], these subgroup specific findings might help elucidate the molecular mechanisms underlying not only the development but also the progression of gastric cancer subtypes.

To provide researchers with extensive information on the most promising genetic predisposition loci so far identified, the present work highlights that - despite the large number of subjects tested in hundreds of studies (including three GWAS [7-9]) - much work is still to be done before we can put together the pieces of the puzzle depicting the relationship between genetic variation and gastric cancer development. For instance, significant associations with intermediate or low level of summary evidence (n=110, **Supplementary Table 2**) should be further investigated with special regard to the assessment of sources of between-study heterogeneity (the main cause of evidence unreliability) as this might ultimately lead to the identification of subgroups of patients for whom the variants do play a role in their disease. Amongst the 335 non-significant associations, high or intermediate level of summary evidence was found only for 26 contrasts (7.7%) (**Table 2**): given that low statistical power and between-study heterogeneity were the main causes of low quality evidence, there is clearly room for assessing whether increasing the sample size and performing subgroup analyses may improve the quality of the summary evidence for the remaining 309 contrasts. Moreover, we could not perform any meta-analysis for 2,685 variants for which fewer than three independent datasets were available. As new data are published, these gaps will hopefully be filled possibly through international collaborative initiatives exploiting online data sharing systems like the one we set up at the above mentioned website.

Only a tiny fraction of the millions of variants hosted by the human genome have been so far investigated, all of them being potentially relevant to gastric carcinogenesis. Recently implemented high-throughput technology has enabled investigators to study hundreds of thousands of SNPs at a time, and some of the best hits (PSCA rs2294008, PSCA rs2976392, PRKAA1 rs13361707 and PLCE1 rs2274223) highlighted by our meta-analysis do derive from GWAS. However, there is still no consensus on the best approach (hypothesis-driven versus hypothesis-free) to study genotype-phenotype associations [10, 11]. Overall, the results of the GWAS so far conducted in the oncology field are not in line with those from previous gene-centric studies (focusing on variants of genes known to have a potential role in carcinogenesis) as they have generated new hypotheses rather than confirming associations previously discovered in candidate gene studies. In fact, the three GWAS carried out in the field of gastric cancer research have suggested four highly significant susceptibility loci (PSCA at 8q24.2 [7], PLCE1 at 10q23 [8], PTGER4/PRKAA1 at 5p13.1 [9], and ZBTB20 at 3q13.31 [9]) never linked before to this disease, whereas only one GWAS [12] confirmed the importance of a locus already pointed out by previous studies (MUC1 at 1q22).

Finally, we looked at variants never tested so far in gastric cancer studies, but already demonstrated to be associated with the risk of other cancers with a high level of cumulative evidence. For example, there were 29 high-quality biomarkers identified in three synopses on breast cancer [13], colorectal cancer [14] and cutaneous melanoma [15] (**Supplementary Figure 8**) and among these, data on gastric cancer are lacking in 20 cases, whereas a meta-analysis of four studies was feasible only for one variant (CTLA4 rs231775).

References

1. Little J, Higgins J. The HuGENet™ HuGE Review Handbook, version 1.0, 2006.

2. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53**:1119-29.

3. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;**25**:3443-57.

4. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials* 2007;**4**:245-53.

5. Crew KD, Neugut AI. Epidemiology of gastric cancer. *World J Gastroenterol* 2006;**12**:354-62.

6. de Martel C, Forman D, Plummer M. Gastric cancer: epidemiology and risk factors. *Gastroenterol Clin North Am* 2013;**42**:219-40.

7. Sakamoto H, Yoshimura K, Saeki N, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* 2008;**40**:730-40.

8. Abnet CC, Freedman ND, Hu N, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* 2010;**42**:764-7.

9. Shi Y, Hu Z, Wu C, et al. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet* 2011;**43**:1215-8.

10. Hoh J, Ott J. Genetic dissection of diseases: design and methods. *Curr Opin Genet Dev* 2004;**14**:229-32.

11. Jorgenson E, Witte JS. A gene-centric approach to genome-wide association studies. *Nat Rev Genet* 2006;**7**:885-91.

12. Saeki N, Saito A, Choi IJ, et al. A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology* 2011;**140**:892-902.

13. Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol* 2011;**12**:477-88.

14. Theodoratou E, Montazeri Z, Hawken S, et al. Systematic meta-analyses and field synopsis of genetic association studies in colorectal cancer. *J Natl Cancer Inst* 2012;**104**:1433-57.

15. Chatzinasiou F, Lill CM, Kypreou K, et al. Comprehensive field synopsis and systematic meta-analyses of genetic association studies in cutaneous melanoma. *J Natl Cancer Inst* 2011;**103**:1227-35.