MINI-REVIEW

Meta-Analysis in Cancer Genetics

Noel A Pabalan

Abstract

Genetic association studies report potentially conflicting findings which meta-analysis seeks to quantify and objectively summarize. Attributing cancer to a single gene variant requires large sample sizes, which may strain resources in a primary study. Properly used, meta-analysis is a powerful tool for resolving discrepancies in genetic association studies given the exponential increase in sample sizes when data are combined. The several steps involved in this methodology require careful attention to critical issues in meta-analysis, heterogeneity and publication bias, evaluation of which can be graphical or statistical. Overall summary effects of a meta-analysis may or may not reflect similar associations when the component studies are sub grouped. Overall associations and that of the subgroups are evaluated for tenability using sensitivity analysis. The low association between a polymorphism and cancer is offset by detectable changes in cancer incidence in the general population making them an important issue from a public health point of view. Asian meta-analytic publications in cancer genetics come from six countries with an output that number from one to two. The exception is China, whose publication output has increased exponentially since 2008.

Key words: Meta-analysis - cancer - genetics

Asian Pacific J Cancer Prev, **11**, 33-38

Introduction

Cancer has a complex hereditary component. With more than a million common variants in the human genome (Sachidanandam et al., 2001), identifying those that are responsible for cancer is a huge undertaking. Just the combination of polymorphisms and cancer alone, there are billions of genetic association analyses that could be performed. Translating to a potentially prodigious number of cancer genetic association publications that do not necessarily agree with each other, meta-analysts have a lot of work to do. Still the relative paucity of meta-analyses compared to the flood of genetic association studies, gives pause to contemplate the direction of this type of research in the Asian setting. In terms of quantity, mainland China is the current world leader in meta-analytical output in cancer genetics. This review is directed at laboratory investigators and clinicians for a better understanding of the state of meta-analysis research in cancer genetics. This review starts with an overview of genetic association studies and meta-analysis, ending with the state of cancer genetics in the Philippines and current meta-analytical output in Asia.

Genetic Association Studies

Genetic association studies evaluate correlations between disease status and genetic variants in a population (Cardon and Bell, 2001). Such studies are more powerful than other approaches (such as linkage) for investigating complex, chronic diseases.

They are the key to explaining relationships between candidate genetic risk factors and complex polygenic diseases. For such diseases, the risk alleles elicit less certainty and more probability, that is presence of a high-risk allele only mildly increases the chance of disease (Hirschhorn et al., 2002). Although, these studies offer a potentially powerful approach in identifying genetic variants that influence susceptibility to common disease (Risch and Merikangas, 1996), the impression persists that they are not consistently replicated by subsequent research (Ioannidis et al., 2001). This lack of reproducibility stems from a number of causes, related to study design, sample size and power issues, as well as true variability between populations (Colhoun et al., 2003).

Nevertheless, research results continue to accumulate, making it increasingly difficult to understand what they tell us and discern the knowledge in this flood of information. No two studies are exactly alike even if they address the same issue. It is clear that number of reports does not indicate confidence in the results: each convincing association is followed by an equally convincing rebuttal (Munafo and Flint, 2004). Conflicting findings may not be easy to reconcile and abundance of primary studies often hinders meaningful integration of results using traditional methods (narrative).

What is Meta-Analysis?

Meta-analysis is an objective, quantitative mode of summarizing research findings and provides the opportunity to help identify genuine associations. Considered at the top of the hierarchy of evidence (Yuan

*College of Natural Sciences, Saint Louis University, Baguio, Philippines *For correspondence: npabalan@alumni.yorku.ca, npabalan@slu.edu.ph*

and Hunt, 2009), this statistical methodology integrates results of independent but related studies to synthesize summaries. It explains inconsistencies as well as moderators and mediators allowing researchers to arrive at conclusions that are more accurate and credible than can be presented in any one primary study or in a narrative review. Being most useful when individual studies are too small to yield valid conclusions, meta-analysis increases power, reduces risk of error and facilitates exploratory analysis which generate hypotheses for future research (Gotzsche, 2000).

Methodology of Meta-Analysis

Literature Search and Data Abstraction

A good meta-analysis starts with a well-formulated and answerable question considering the time, cost and available resources. In gene-disease studies, this includes availability of a good number of primary studies that address associations of key genetic polymorphisms (synonymous or non-synonymous single nucleotide as well as repeats) with chronic disease.

Search strategies typically involve electronic retrieval of all available literature which in genetic association includes digital sources such as PubMed and Medline. Used exclusively, however, such searches may miss a substantial proportion of relevant studies (Bai et al., 2007). For greater precision in this step, additional measures to exhaustively identify eligible studies include manual searching of relevant journals, references lists and personal contact with researchers.

Eligibility criteria, which relate to quality and combinability of associations, subjects and outcomes, are defined a priori. The next step is abstraction of both qualitative (e.g. population characteristics) and quantitative (e.g. sample sizes and genotypic frequencies) data from the collection of eligible studies.

Summary Effects Calculations

In meta-analyses of gene-disease association studies, summary effects are presented under various genetic models so that dominant, codominant and recessive patterns are elucidated (Minelli et al., 2005). The abstracted quantitative data is used to calculate a summary effect (e.g. odds ratios [OR] and 95% confidence intervals [CI]). The OR has convenient mathematical properties, which allow for ease in combining data and testing the overall effect for significance (Egger et al., 1997). Results of each study are graphically presented in a forest plot (Figure 1). In this example, ORs of 11 studies, each represented by a black square and a horizontal line, representing the point estimate and 95% CIs. The solid vertical line (labeled 1 on the x-axis) corresponds to null effects. One observes that the CIs of all the studies cross this vertical line, indicating that the effect estimates were non-significant ($p > 0.05$). The area of the black squares reflects the weight of the study in the meta-analysis, the larger the area, the more weight the study contributes to the pooled OR (column 4 of Figure 1).

The diamond at the bottom of the forest plot represents the pooled summary effect (OR 1.21, 95% 1.09-1.35) calculated using a fixed-effects model, which shows that the catechol-O-methyl transferase (COMT) genotype (GG vs. AA comparison) is significantly associated with breast

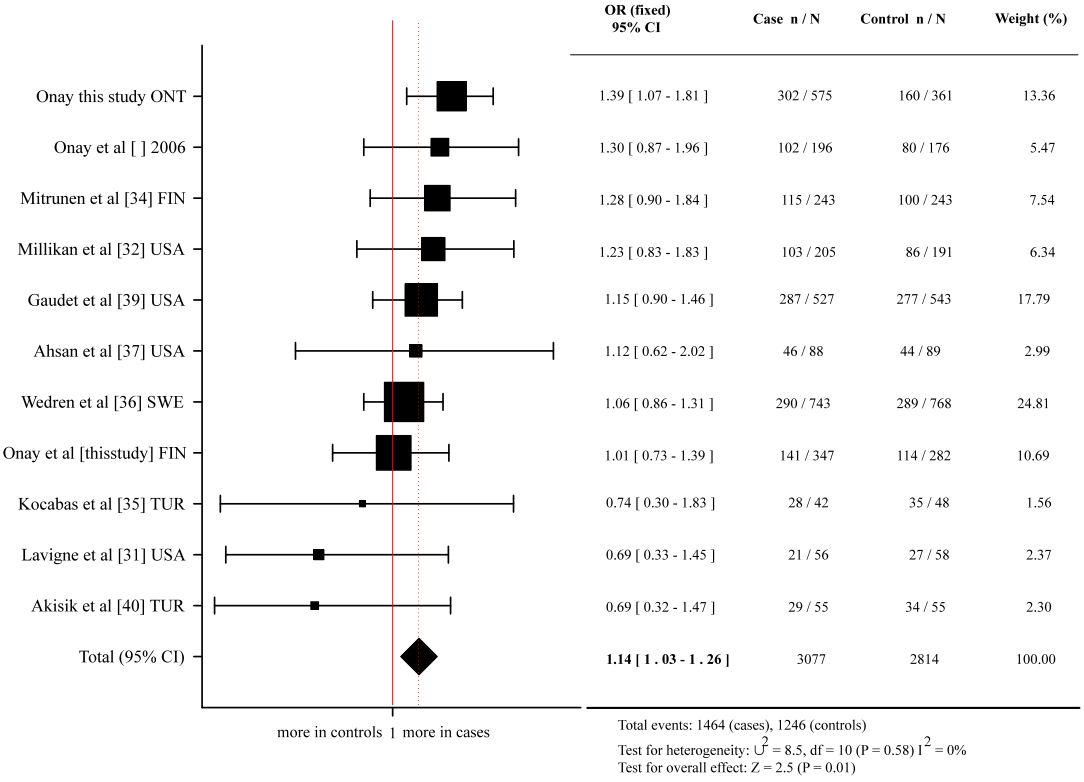

Figure 1. Meta-analysis of Risk Estimates (OR) of COMT Low Enzyme Activity (Methionine) Allele in Case Control Studies of Caucasian Breast Cancer Samples and Population Controls. For each study, the OR estimate and its 95% CI was plotted with a box and horizontal line. ® Indicates pooled OR and its 95% CI (Onay et al., 2008)

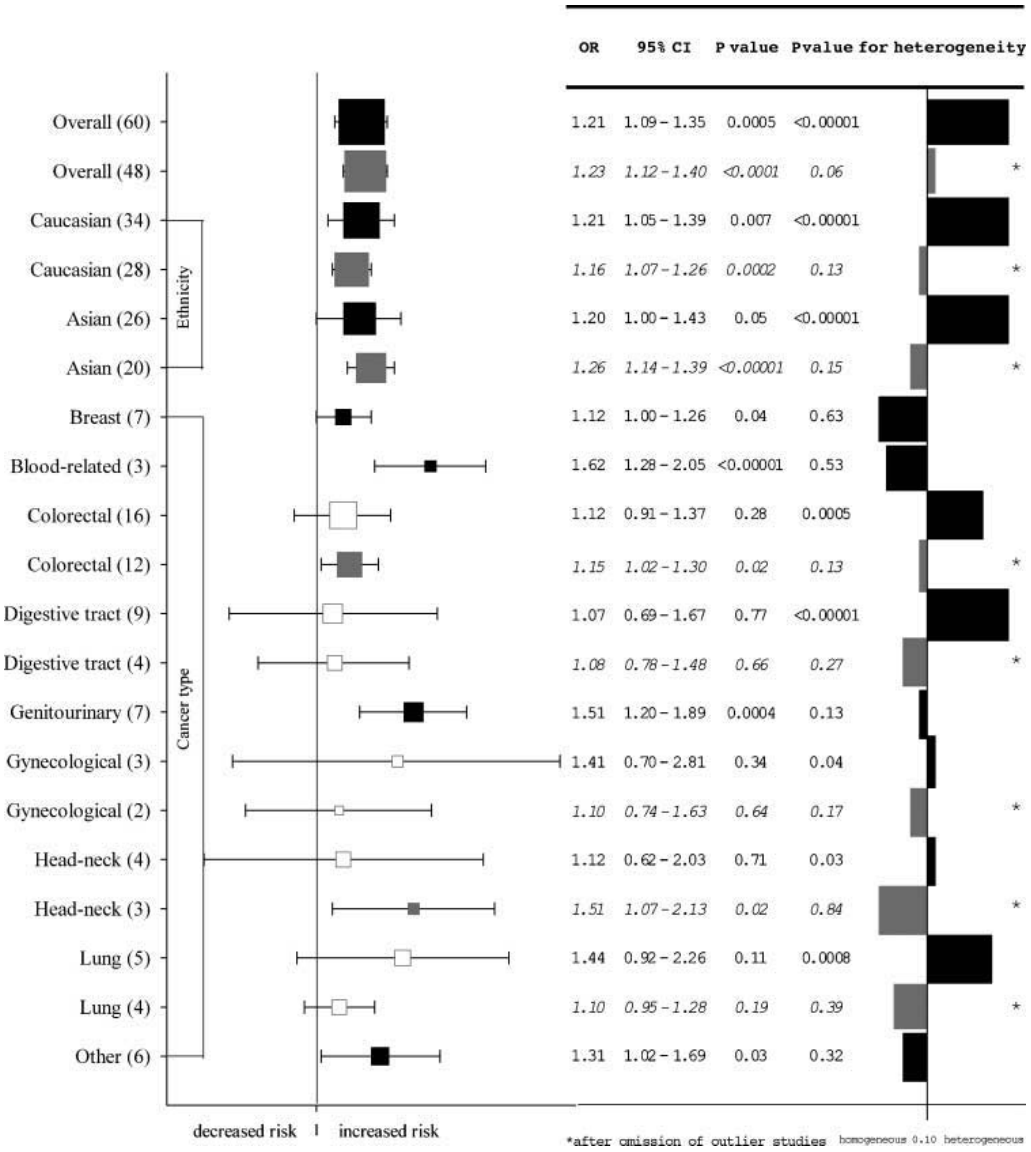

Figure 2. Forest and Heterogeneity Plots of Pooled Effects of the CCND1 G870A Polymorphism in AA Homozygotes. For each study, the OR estimate and its 95% CI was plotted with a box and horizontal line. Numbers in parentheses along the Y-axis indicate number of studies. Squares, summary estimates; n, significance; n, nonsignificance Larger squares: higher sample sizes. Lines on either side of squares: 95% CI. Black bars, main effects on heterogeneity with outlier studies; gray bars, effects of removal of outliers indicated by asterisks (*). P values were set at 0.05 for OR effects and <0.10 for heterogeneity (Pabalan et al., 2008)

cancer risk (P = 0.01). The dotted line is plotted vertically through the diamond, which crosses the horizontal lines (and squares) in all individual studies. This indicates a fairly homogeneous set of component studies. Indeed, the test for heterogeneity gives a non-significant P value of 0.58 (Onay et al., 2008).

Sensitivity and Subgroup Analyses

The summary or pooled effects need to be tested further to ensure rigor of this methodology. The tests include sensitivity analysis to determine robustness of pooled effects. In Figure 2, the forest plot was modified to indicate results of sensitivity analysis, where detected outliers (by means of the Galbraith plot (Figure 3)) were omitted from the analysis with subsequent recalculation of the summary effects. In Figure 2, sensitivity analysis narrowed all the confidence intervals of the overall summary effects, ethnic groups and cancer types conferring greater precision on the pooled ORs. Sensitivity analysis did not affect the overall ORs and that of the ethnic groups as much as it did on the cancer types suggesting that the overall and ethnic ORs are more robust than those of the cancer types.

Subgroup analysis allows identification of association in key population groups using various parameters such as ethnicity and geography. The forest plot in Figure 2 shows on the Y axis the 60 component studies in the meta-analysis sub grouped by ethnicity and cancer type. Pooled effects in the ethnic subgroups did not differ much from the overall summary effects in terms of increased risk and statistical significance (Pabalan et al., 2008).

Issues in Meta-Analysis

Heterogeneity

Component studies in a meta-analysis may be evaluated on the basis of their similarity to each other

and applying the fixed-effects method of analysis (Mantel and Haenszel.361959) depends on the assumption that associations are the same across studies and recognizing that the collection of eligible literature is not heterogeneous. Heterogeneity is the methodological, epidemiological and clinical dissimilarity across various studies and meta-analysts spend considerable effort in addressing this issue. In Figure 1, the p value in the test for heterogeneity is highly significant (P<0.0001), necessitating an adjustment with the random effects analysis model (DerSimonian and Laird, 1986) which assumes variability across populations usually resulting in a wider CI (Egger et al., 1997). Graphically, plots are used to detect (Galbraith, 1988), and summarize heterogeneity (Figure 3). Figure 2 identified a total of 12 studies above and below the 95% confidence limits to be outliers. Exclusion of these outliers impacted upon heterogeneity of the original findings and is summarized in Figure 3 (Pabalan et al., 2008).

Statistically, heterogeneity is estimated using a chi-square-based Q test (Zintzaras and Ioannidis, 2005) and quantified with the I2 metric which shows what proportion of the total variation across studies is beyond chance (Higgins and Thompson, 2002).

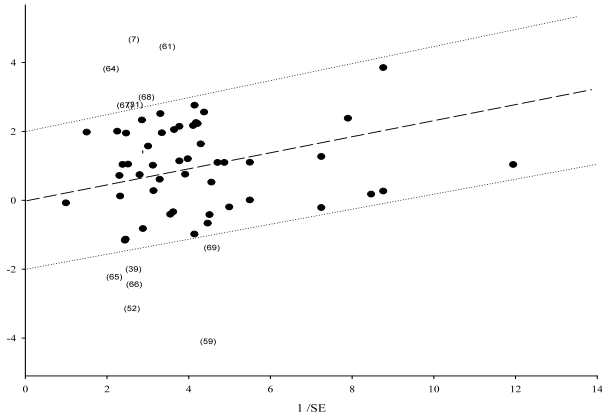

Figure 3. Galbraith Plot Analysis to Evaluate Heterogeneity. For each point, the ratio of the log odds ratio to its SE is plotted against the reciprocal of the SE. Less precise outcomes appear toward the left of the graph and the largest studies appear toward the right. The dotted lines positioned two units above and below the solid line delimit the area, which, in the absence of heterogeneity, 95% of the points would be expected to lie outside. Reference numbers identify the studies that lie outside the 95% confidence limits (Pabalan et al., 2008)

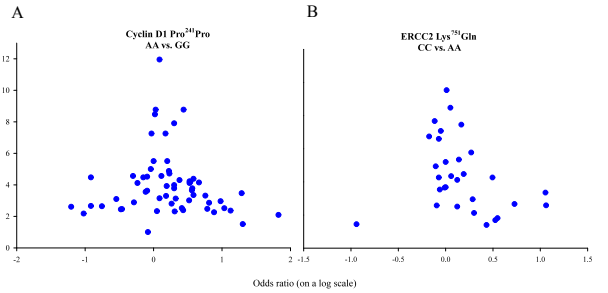

Figure 4. Funnel Plot Analysis to Detect Publication Bias. Each point represents a separate study for the indicated association. For each study, the OR is plotted on a logarithmic scale on the X axis against the precision (1 / SE) on the Y axis (Pabalan et al., 2008; Pabalan, 2010)

Publication Bias

Publication bias (Stroup et al., 2000) is an issue where significant findings receive priority in published literature over those that fail to reject the null hypotheses resulting in non-significant associations. Graphically, publication bias is evaluated with the funnel plot (Figures 4A and B) where effect estimate from each study in the meta-analysis is scattered against a measure of its precision usually 1/SE (standard error). Figure 4A shows a symmetrical distribution of the points with small studies scattered along the length of the X axis but still centered around the OR estimates from large, more precise studies indicating absence of bias (Pabalan et al., 2008). Figure 4B shows an asymmetrical distribution of the points with small studies concentrated on one side of the plot indicating presence of publication bias (Pabalan, 2010). Given the subjective aspect to interpreting funnel plots, publication bias is formally evaluated with either the Egger's regression asymmetry and/or the Begg's and Mazumdar's rank correlation tests.

Value of Meta-Analysis in Cancer Genetics

Interest in genetic predisposition to common disease has grown over the past decade (Sagoo et al., 2009). Rapid advancements in high-throughput genotyping have resulted in large amounts of published epidemiological evidence on gene-disease associations. Genetic polymorphisms have long been established to play a role in cancer susceptibility (Gonzalez, 1995) where they are involved in several mechanisms such as DNA repair regulated by specific genes and cell cycling. The role of key polymorphisms in genes involved in these mechanisms has been meta-analyzed (Pabalan et al., 2008; Pabalan, 2010). Such undertakings increase statistical power enough for information to be useful for public health advice in cancer risk. Information about one's predisposition to cancer may elicit modest behavioral changes (quit smoking, reduce alcohol consumption adopt dietary changes or avoid particular occupational exposures) which on the population level are likely to make a large public health impact (Brennan, 2002).

Although association between the polymorphisms and cancer is low, with odds ratios ranging between 1.1 and 1.5 (Zintzaras and Lau, 2008), their diffusion in the general population make them an important issue from a public health point of view. Even a small mathematical change in an association between a polymorphism and cancer could result in a detectable change in the calculation of cancer incidence in the general population (Taioli, 2005). The significance of meta-analysis produces two outcomes: (i) facilitate critical transfer of knowledge from researcher to clinician, allowing for re-evaluation and identification of high-risk subjects because of their genetic make-up (Taioli, 2005), (ii) enable analyses of important patient subgroups and identify those individuals at high risk for cancer (Ioannidis, 2004).

To increase the probability of attributing cancer to a single gene variant from a tiny susceptibility locus (Lohmueller et al., 2003), large sample sizes are needed. For a typical odds ratio of 1.3, where marker and disease

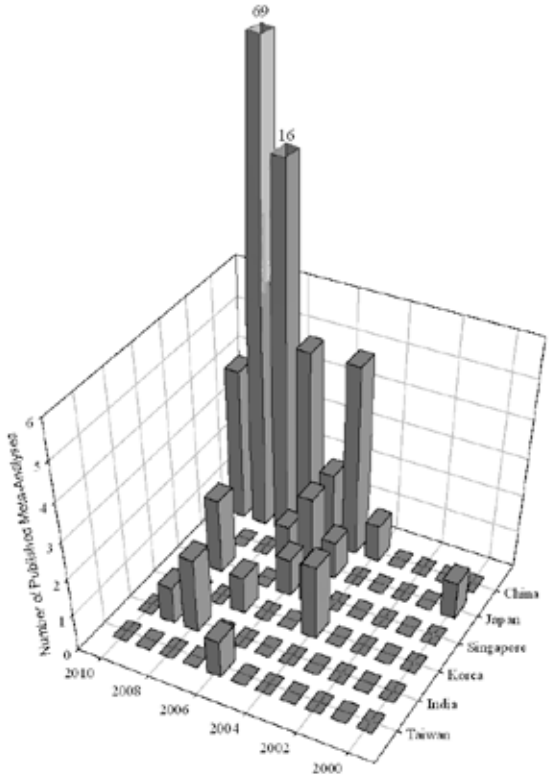

Figure 5. Asian Outputs of Meta-Analyses in Cancer Genetics. Numbers of published meta-analyses in mainland China for 2008 and 2009 were 16 and 69, respectively.

allele frequencies match, sample sizes of 2,000 to 10,000 cases and controls are required to obtain satisfactory power (80%) (Zintzaras and Lau, 2008; Zondervan and Cardon, 2004). Although a single laboratory or trial might not be able to obtain such numbers, the combined world literature might if there was a way to analyze the data jointly. Meta-analysis is an attractive and cheaper alternative to the primary study, which when large, is bound to be expensive and logistically problematic. Done rigorously, meta-analysis reaches the necessary number of subjects, and effects can be detected with more confidence and statistical precision.

Growing interest in gene-disease associations has produced large amounts of published epidemiological evidence. Nowhere is this productivity more evident in cancer research where genetic predisposition is characteristic of the etiology of this disease.

Cancer Genetics in the Philippines

In the Philippines, cancer is one of ten leading causes of death (DOH, 2004). A PubMed search of the associations of gene polymorphisms with this disease yielded two case-control studies, one based in UP-PGH (Alcasabas et al., 2008), and the other based at the University of Toronto in Canada (Liede et al., 2003) . Filipino participants in the former study were children and women in the latter. For both studies, the probability of detecting an association assuming a genotypic risk of 1.3 was 31%. Clearly, higher sample sizes are needed to raise the probability to an acceptable level of 80%. A meta-analysis incorporating these studies would accommodate this power increase and include the Filipino ethnicity into a larger subset of subgroup analysis where various risks of other ethnic groups are compared in a true epidemiological fashion.

Although polymorphisms in susceptibility genes are diffuse in the healthy population, their absolute frequency varies with ethnicity (Taioli, 2005). For example, the Val allele of the 655 polymorphism in the HER-2 gene, alterations of which are implicated in breast cancer, accounted for 9% in the Filipino population and 20% in the Caucasian population (Ameyaw et al., 2002). Another polymorphism, the polyglutamine repeat of exon 1 in the androgen receptor (AR) gene has been proposed to be a modifier of breast cancer risk (Giguere et al., 2001). A study on the association of breast cancer and the trinucleotide repeat polymorphism (CAGn) in this gene was conducted with the conclusion of an observed odds ratio of 0.47 (95% confidence interval 0.28–0.8) suggesting significant protection for Filipino women with ≤25 units of the CAG repeat allele (Liede et al., 2003). Complementing these polymorphic profiles with other functional studies on CAG repeats (Ozcelik, 2005) provides a better understanding of the genetic aspects of breast cancer etiology.

Asian Output of Meta-Analyses in Cancer Genetics

A PubMed Search was performed using the search terms, "meta-analysis", "cancer", polymorphism" and appended the Asian country in sequence. Figure 5 shows results of the search wherein six Asian countries published meta-analyses in this field from 2000 to 2010. The early part of the 21st century (2000 to 2003) showed minimal meta-analyses output. From 2004 to 2009, meta-analyses publications among the Asian countries excepting China numbered from one to two. From 2005 to 2007, the People's Republic of China published a total of 12 meta-analyses. This three-year output was superseded the following year (2008) with 16 publications. In 2009, meta-analyses output from this country increased more than four-fold to 71 publications. The year 2010 has barely started and China has already published four meta-analyses. One can appreciate the breadth of this country's capability to utilize output from genotyping technology in cancer research. Such outputs probably place genetic profiling closer to its incorporation and translation into clinical intervention.

Acknowledgements

The Philippine Department of Science and Technology (DOST) awarded Noel Pabalan with a Balik-Scientist Status.

References

Alcasabas P, Ravindranath Y, Goyette G, et al (2008). 5,10-methylenetetrahydrofolate reductase (MTHFR) polymorphisms and the risk of acute lymphoblastic leukemia (ALL) in Filipino children. *Pediatr Blood Cancer*, **51**, 178-82.

Ameyaw MM, Tayeb M, Thornton N, et al (2002). Ethnic variation in the HER-2 codon 655 genetic polymorphism

previously associated with breast cancer. *J Hum Genet*, **47**, 172-5.
Bai Y, Gao J, Zou D, Li Z (2007). Is MEDLINE alone enough for a meta-analysis? *Aliment Pharmacol Ther*, **26**, 125-6; author reply 126.
Brennan P (2002). Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* , 381-7.
Cardon LR, Bell JI (2001). Association study designs for complex diseases. *Nat Rev Genet*, **2**, 91-9.
Colhoun HM, McKeigue PM, Davey Smith G (2003). Problems of reporting genetic associations with complex outcomes. *Lancet*, **361**, 865-72.
DerSimonian R, Laird N (1986). Meta-analysis in clinical trials. *Control Clin Trials*, **7**, 177-88.
Department of Health (2004). Philippine Health Statistics.
Egger M, Smith GD, Phillips AN (1997). Meta-analysis: principles and procedures. *BMJ*, **315**, 1533-7.
Galbraith RF (1988). A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med*, **7**, 889-94.
Giguere Y, Dewailly E, Brisson J, et al (2001). Short polyglutamine tracts in the androgen receptor are protective against breast cancer in the general population. *Cancer Res*, **61**, 5869-74.
Gonzalez FJ (1995). Genetic polymorphism and cancer susceptibility: 14th Sapporo Cancer Seminar. *Cancer Res*, **55**, 710-5.
Gotzsche PC (2000). Why we need a broad perspective on meta-analysis. It may be crucially important for patients. *BMJ*, **321**, 585-6.
Higgins JP, Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med*, **21**, 1539-58.
Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002). A comprehensive review of genetic association studies. *Genet Med*, **4**, 45-61.
Ioannidis JP (2004). The value of meta-analysis in rheumatology research. *Autoimmun Rev*, **3 Suppl 1**, S57-9.
Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001). Replication validity of genetic association studies. *Nat Genet*, **29**, 306-9.
Liede A, Zhang W, De Leon Matsuda ML, Tan A, Narod SA (2003). Androgen receptor gene polymorphism and breast cancer susceptibility in The Philippines. *Cancer Epidemiol Biomarkers Prev*, **12**, 848-52.
Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet*, **33**, 177-82.
Mantel N, Haenszel W (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*, **22**, 719-48.
Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J (2005). The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol*, **34**, 1319-28.
Munafo MR, Flint J (2004). Meta-analysis of genetic association studies. *Trends Genet*, **20**, 439-44.
Onay UV, Aaltonen K, Briollais L, et al (2008). Combined effect of CCND1 and COMT polymorphisms and increased breast cancer risk. *BMC Cancer*, **8**, 6.
Ozcelik H, Jarjanazi, H., Pabalan, N (2005). Cloning and characterization of expanded CAG-repeat containing sequence(s): identification of candidate breast cancer predisposition genes. *Genetic Engineering Molec Biol*, **41**, A964244.
Pabalan N, Bapat B, Sung L, et al (2008). Cyclin D1 Pro241Pro (CCND1-G870A) polymorphism is associated with increased cancer risk in human populations: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*, **17**, 2773-81.
Pabalan N, Francisco-Pabalan O, Sung L, Jarjanazi H, Ozcelik H (2010). Meta-analysis of two polymorphisms, Lys751Gln and Asp312Asn, in the DNA repair gene, XPD (ERCC2) with breast cancer. *Breast Cancer Res Treat*, **17**, 27
Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. *Science*, 273, 1516-7.
Sachidanandam R, Weissman D, Schmidt SC, et al (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**, 928-33.
Sagoo GS, Little J, Higgins JP (2009). Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLoS Med*, **6**, e28.
Stroup DF, Berlin JA, Morton SC, et al (2000). Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*, **283**, 2008-12.
Taioli E (2005). Biomarkers of genetic susceptibility to cancer: applications to epidemiological studies. *Future Oncol*, **1**, 51-6.
Yuan Y, Hunt RH (2009). Systematic reviews: the good, the bad, and the ugly. *Am J Gastroenterol*, **104**, 1086-92.
Zintzaras E, Ioannidis JP (2005). Heterogeneity testing in meta-analysis of genome searches. *Genet Epidemiol*, **28**, 123-37.
Zintzaras E, Lau J (2008). Trends in meta-analysis of genetic association studies. *J Hum Genet*, **53**, 1-9.
Zondervan KT, Cardon LR (2004). The complex interplay among factors that influence allelic association. *Nat Rev Genet*, **5**, 89-100.

Personal Profile: Noel Pabalan

Noel Pabalan received his Ph.D. in Biology from York University (Toronto, Canada) in 1998. From his postdoctoral work in cardiovascular genetics from the University of Toronto from 1999 to 2000, his subsequent experiences were in molecular biology of various chronic diseases ranging from osteoarthritis to cancer genetics. His current research is in meta-analyses of various polymorphisms that may be associated with cancer.