Selection of genes for gene-environment interaction studies: a candidate pathway-based strategy using asthma as an example.

Marta Rava, Ismaïl Ahmed, Florence Demenais, Margaux Sanchez, Pascale Tubert-Bitter, Rachel Nadif

▶ **To cite this version:**

Marta Rava, Ismaïl Ahmed, Florence Demenais, Margaux Sanchez, Pascale Tubert-Bitter, et al.. Selection of genes for gene-environment interaction studies: a candidate pathway-based strategy using asthma as an example.. Environmental Health, 2013, 12 (1), pp.56. 10.1186/1476-069X-12-56 . inserm-00843444

HAL Id: inserm-00843444

https://inserm.hal.science/inserm-00843444

Submitted on 11 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ENVIRONMENTAL HEALTH

HYPOTHESIS Open Access

Selection of genes for gene-environment interaction studies: a candidate pathway-based strategy using asthma as an example

Marta Rava[1,2*†], Ismaïl Ahmed[2,3†], Florence Demenais[4,5], Margaux Sanchez[1,2], Pascale Tubert-Bitter[2,3] and Rachel Nadif[1,2]

Abstract

Background: The identification of gene by environment (GxE) interactions has emerged as a challenging but essential task to fully understand the complex mechanism underlying multifactorial diseases. Until now, GxE interactions have been investigated by candidate approaches examining a small number of genes, or agnostically at the genome wide level.

Presentation of the hypothesis: In this paper, we propose a gene selection strategy for investigation of gene-environment interactions. This strategy integrates the information on biological processes shared by genes, the canonical pathways to which they belong and the biological knowledge related to the environment in the gene selection process. It relies on both bioinformatics resources and biological expertise.

Testing the hypothesis: We illustrate our strategy by considering asthma, tobacco smoke as the environmental exposure, and genes sharing the same biological function of "response to oxidative stress". Our filtering strategy leads to a list of 28 pathways involving 182 genes for further GxE investigation.

Implications of the hypothesis: By integrating the environment into the gene selection process, we expect that our strategy will improve the ability to identify the joint effects and interactions of environmental and genetic factors in disease.

Keywords: Gene by environment interactions, Oxidative stress, Smoking, Pathway-based gene selection

Background

Until recently, gene by environment (GxE) interaction studies were performed by means of candidate approaches including only a small number of genes. Gene selection in candidate studies relies on 1) known functions of gene sets sharing biological processes, and/or functionally interacting within biological networks; or 2) the mode of action of the environmental factors through relevant pathways in which genes are involved [1]. With the advent of high-throughput genotyping technologies, GxE interactions are starting to be explored at the genome wide level but this approach involves the following difficulties: 1) the heterogeneity of environmental exposures; 2) the "agnostic" nature of the genome-wide approach, which does not make use of prior knowledge on biological processes and/or pathways; and 3) the requirement of stringent thresholds to declare an GxE interaction significant because of the very large number of statistical tests conducted [2].

In this scenario, the classical candidate gene approach can be extended to the selection of large sets of genes. In this paper, we propose a strategy for obtaining a large gene set that integrates the information on biological processes shared by genes, the canonical pathways to which they belong and the biological knowledge related to the environmental exposure studied in the gene selection process.

* Correspondence: marta.rava@inserm.fr
†Equal contributors
[1]Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Respiratory and Environmental Epidemiology Team, F-94807, Paris, Villejuif, France
[2]University Paris-Sud, UMRS 1018, F-94807, Paris, Villejuif, France
Full list of author information is available at the end of the article

BioMed Central

© 2013 Rava et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The asthma example

Asthma is a complex heterogeneous multifactorial disorder resulting from genetic and environmental factors [3] and whose etiology remains poorly understood. The increase in asthma prevalence in recent decades has led to extensive research regarding the environmental determinants that may have changed over the last 30 years. There have also been considerable efforts to characterize the genetic determinants of asthma, including candidate gene studies, genome-wide linkage screens followed by positional cloning studies and more recently genome-wide association studies (GWAS) [4]. Although these studies have been successful in identifying novel loci, the genetic factors identified explain only a small part of the genetic component of asthma. One of the reasons is that many genetic factors are likely to be involved in the development, the activity and the severity of asthma. Furthermore, they act primarily through complex mechanisms that involve interactions with environmental factors, or with other genes through pathways or networks. The effect of such genetic factors may be missed if their interactions with the environment are not taken into account, or if genes are considered alone, regardless of the biological functions they shared or the pathways they are involved in [5]. Overall, understanding the mechanisms through which genes and the environment interact represents one of the major challenges for pulmonary researchers. The first Genome-Wide Environment Interaction Study (GWEIS) in asthma [6] identified no statistically significant interaction at the genome-wide level, not even with Single Nucleotide Polymorphisms (SNPs), which were shown to interact with the environment in previous candidate studies.

In response to environmental exposures, adaptive responses for protection against environmental toxic insults are activated through metabolic pathways. Among the several metabolic pathways that could be investigated in asthma, the response to oxidative stress is of major interest: the amount of biological evidence of the role of oxidative stress in asthma is increasing [7], and tobacco smoke is related to oxidative stress. Tobacco smoke is also a risk factor for asthma. Active smoking has been found to be associated with the incidence of asthma during adolescence in a dose-dependent manner [8] and with asthma severity in asthmatic cases [9]. Regular smoking was associated with increased risk of new-onset asthma among adolescents in a prospective cohort study [10], and active smoking has a deleterious role on asthma [11]. To our knowledge, only one study focused on gene by smoking interactions on asthma in adults by considering 18 key genes involved in the same pathway: the metabolism of xenobiotics. Some of these genes were also involved in the response to oxidative stress, and SNPs in seven of them were significantly associated with the risk of asthma in adult smokers or non-smokers [12].

Presentation of the hypothesis

In this paper, we propose a strategy for selecting genes to be investigated in GxE interaction studies. This strategy involves the information on biological processes shared by the genes, the canonical pathways to which they belong to and biological knowledge related to the environment into the gene selection process. We hypothesize that this strategy will provide an expanded and enriched biologically plausible list of candidate genes for further GxE studies.

This strategy follows three successive steps (see Figure 1): 1) step 1 (gene selection): selection of a set of genes sharing a biological process known to be related with the outcome or the disease of interest, 2) step 2 (pathway enrichment): selection of physically and/or chemically related gene pathways that are enriched in genes belonging to the gene set selected in step 1. Among the pathways that constitute a biological process, we considered the signaling and/or

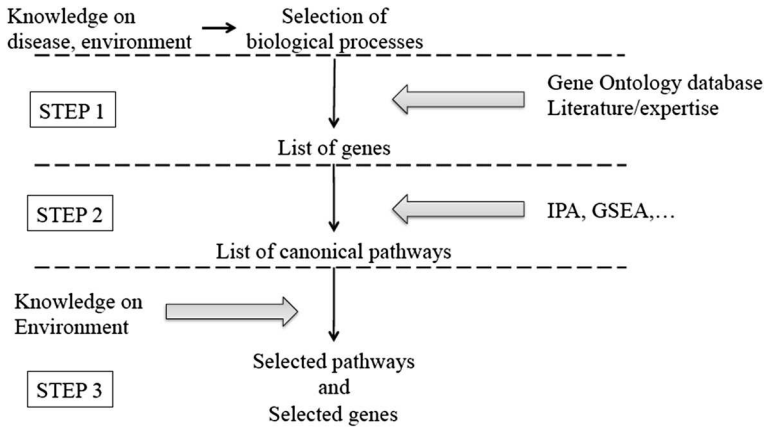

Figure 1 The three-step strategy.

metabolic pathways, also known as canonical pathways, which better suit the subsequent environmental integration step, and 3) step 3 (environment integration): selection of canonical pathways known to be potentially related to the environmental factor of interest among the pathways selected in step 2. The final set of genes includes the genes selected in step 1 that belong to the canonical pathways selected in step 3. Note that step 3 critically relies on the user's own expertise.

Testing the hypothesis

To illustrate our strategy, we consider asthma exposure to tobacco smoke as the environmental factor, and the genes involved in the response to oxidative stress.

Step 1 (gene selection)

The set of genes was obtained from the Gene Ontology (GO) database (Gene Ontology Consortium [13,14]), as described in the online tutorial [see Additional file 1]. The GO project is a bioinformatics initiative that aims at standardizing the representation of genes and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data, as well as tools to access and process this data. We used the term "*response to oxidative stress*" (GO:0006979) which encompasses gene products that are involved in any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of oxidative stress, a state often resulting from exposure to high levels of reactive oxygen species, e.g. superoxide anions, hydrogen peroxide, and hydroxyl radicals. We obtained a set of 387 genes, including all genes previously investigated in candidate GxE interaction studies in respiratory epidemiology such as *MPO*, *CAT*, *GCLM*, *GCLC*, *GSTP1*, *NQO1* [15-21], and some genes in the study by Polonikov *et al.* [12]. We further enlarged the gene set by using our own expertise, GWAS literature reviews, and biological studies [22-26]. A total of 411 genes were then considered for the next step.

Step 2 (pathway enrichment)

This step consists in identifying canonical pathways that contain a statistically significant excess of genes from the set of 411 genes selected in step 1. This pathway analysis can be conducted by using several tools such as Ingenuity Pathway Analysis (IPA, [27]) or Gene Set Enrichment Analysis (GSEA [28,29]). These software solutions differ in terms of the biological databases they rely on (KEGG, Biocarta, Reactome, Pubmed, STRING…) and the methods used to assess the statistical significance of the pathways.

All gene symbols were recognized by IPA but not by GSEA (390 out of 411). IPA gave 277 canonical pathways that contained at least 5 of the set of 411 genes selected in step 1 and which were significantly enriched in these genes ($p < 0.05$). IPA P-values for pathway enrichment testing were obtained with Fisher's exact tests, with a Benjamini–Hochberg correction for multiple testing determined by the ratio of the number of genes from the gene set to the total number of genes in the pathways from the IPA library. GSEA provided no more than the top 100 canonical pathways ($p < 1.06\ 10^{-12}$). Comparing the results provided by both software packages is difficult as the names of the pathways and the genes involved in them are not standardized. Therefore, we decided to perform the third step with the largest list

Table 1 Distribution of the 182 genes by canonical pathways involved in the tobacco smoke metabolism

Canonical pathways	P-value*	N of genes
NRF2-mediated Oxidative Stress Response	1.58E-31	46
Glutathione Redox Reactions I	1.26E-25	16
Xenobiotic Metabolism Signaling	6.31E-22	44
Aryl Hydrocarbon Receptor Signaling	3.98E-21	32
Mitochondrial Dysfunction	6.31E-19	30
Glutathione-mediated Detoxification	2.00E-18	15
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	3.98E-16	31
Acute Phase Response Signaling	2.51E-14	28
Antioxidant Action of Vitamin C	1.58E-12	20
IL-8 Signaling	1.26E-11	26
Apoptosis Signaling	2.95E-09	16
Superpathway of Citrulline Metabolism	4.07E-08	7
Superoxide Radicals Degradation	1.05E-07	5
IL-6 Signaling	1.74E-07	16
iNOS Signaling	2.00E-07	10
VEGF Signaling	1.29E-06	13
fMLP Signaling in Neutrophils	1.95E-06	14
Chemokine Signaling	3.02E-06	11
VEGF Family Ligand-Receptor Interactions	5.50E-05	10
NF-KB Signaling	6.31E-05	15
CCR5 Signaling in Macrophages	1.10E-04	9
IL-17A Signaling in Airway Cells	3.89E-04	8
Nucleotide Excision Repair Pathway	4.07E-04	6
IL-1 Signaling	1.10E-03	9
Nicotine Degradation II	2.82E-03	6
Nicotine Degradation III	7.08E-03	5
CCR3 Signaling in Eosinophils	1.51E-02	8
eNOS Signaling	2.63E-02	8

*P-values for pathway enrichment testing as calculated by IPA.

of pathways and genes *i.e.* the 277 pathways obtained from IPA.

Step 3 (environment integration)

Based on our own expertise, we selected the canonical pathways identified at step 2 that are involved in tobacco smoke metabolism, thus allowing the step 1-gene set to be filtered. Among the 277 canonical pathways identified in step 2, we selected 28 of them (pathway enrichment *P*-values ranging from $2.63x10^{-2}$ to $1.58x10^{-31}$) [see Additional file 2: Table S1 and Table S2]. These 28 pathways included from 5 up to 47 genes (15–20 genes on average), 61% of them being involved in more than one pathway. Two hundred and twenty-nine genes from the initial set of 411 genes did not map to any of the selected pathways and were dropped, leading to a final set of 182 genes (Table 1).

Implications of the hypothesis

The candidate pathway-based strategy described here was able to select a large number of candidate genes to be tested for interaction with tobacco on asthma. This filtering strategy exploits recent developments in bioinformatics resources that are originally combined with the literature and our own expertise on the metabolism of compounds related to a given environmental factor. This filtering strategy could be applied to other environmental factors related to oxidative stress and asthma, such as outdoor air pollutants or the metabolism of cleaning agents. Together with an expanded and enriched list of candidate genes, the interest of such an approach is also dependent on accurate assessment of environmental exposure. Interestingly, the same list of genes can be used for GxE studies on other diseases characterized by oxidative stress and tobacco smoke, such as lung cancer. By appropriately integrating the knowledge of the environmental factor into the gene selection, we expect that the strategy proposed here will improve the ability to identify the joint effects and interactions of environmental and genetic factors, and will contribute to a better understanding of the etiology of complex diseases.

Additional files

Additional file 1: Tutorial: Tutorial on how to extract genes from Gene Ontology.

Additional file 2: Table S1. List of the 182 genes selected using the pathway-based filtering strategy. **Table S2.** List of the 28 pathways and the relevant genes selected using the pathway-based filtering strategy.

Abbreviations

GxE: Gene by environment; GO: Gene ontology; GSEA: Gene set enrichment analysis; GWAS: Genome-wide association studies; GWEIS: Genome-wide environment interaction study; IPA: Ingenuity pathway analysis; SNP: Single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RN reviewed the literature, designed and developed the strategy, selected the genes and pathways and drafted the manuscript. MR reviewed the literature, participated in the gene-selection process and drafted and revised the manuscript. FD helped to develop the strategy and revised the manuscript critically for important intellectual content. MS participated in data acquisition and revised the manuscript. PTB took part in the development of the strategy and revised critically the manuscript. IA participated in the gene selection process, helped to draft the manuscript and revised critically the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Research funded in part by Agence Nationale de la Recherche (ANR) (ANR- 2010-PRSP-003, and the Large-Scale Genome-Wide Association Study of Asthma (GABRIEL), a multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community (contract 018996 from the European Commission).

Author details

[1]Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Respiratory and Environmental Epidemiology Team, F-94807, Paris, Villejuif, France. [2]University Paris-Sud, UMRS 1018, F-94807, Paris, Villejuif, France. [3]Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Biostatistics Team, F-94807, Paris, Villejuif, France. [4]Inserm, U946, F-75010, Paris, France. [5]Institut Universitaire d'Hématologie, University Paris Diderot, Sorbonne Paris Cité, F-75007, Paris, France.

Received: 3 April 2013 Accepted: 2 July 2013
Published: 3 July 2013

References

1. Kauffmann F, Nadif R: **Candidate gene-environment interactions.** *J Epidemiol Community Health* 2010, **64:**188–189.
2. Ober C, Vercelli D: **Gene-environment interactions in human disease: nuisance or opportunity?** *Trends in genetics: TIG* 2011, **27:**107–115.
3. Von Mutius E: **Gene-environment interactions in asthma.** *J Allergy Clin Immunol* 2009, **123:**3–11.
4. Holloway JW, Yang IA, Holgate ST: **Genetics of allergic disease.** *J Allergy Clin Immunol* 2010, **125**(2 Suppl 2)**:**81–94.
5. Liu C, Maity A, Lin X, Wright RO, Christiani DC: **Design and analysis issues in gene and environment studies.** *Environ Health global access scie source* 2012, **11:**93.
6. Ege MJ, Strachan DP, Cookson WOCM, Moffatt MF, Gut I, Lathrop M, Kabesch M, Genuneit J, Büchele G, Sozanska B, Boznanski A, Cullinan P, Horak E, Bieli C, Braun-Fahrländer C, Heederik D, Von Mutius E: **Gene-environment interaction for childhood asthma and exposure to farming in Central Europe.** *J Allergy Clin Immunol* 2011, **127:**1–4. 138–44, 144.e.
7. Chung KF, Marwick JA: **Molecular mechanisms of oxidative stress in airways and lungs with reference to asthma and chronic obstructive pulmonary disease.** *Ann N Y Acad Sci* 2010, **1203:**85–91.
8. Genuneit J, Weinmayr G, Radon K, Dressel H, Windstetter D, Rzehak P, Vogelberg C, Leupold W, Nowak D, Von Mutius E, Weiland SK: **Smoking and the incidence of asthma during adolescence: results of a large cohort study in Germany.** *Thorax* 2006, **61:**572–578.
9. Siroux V, Pin I, Oryszczyn MP, Le Moual N, Kauffmann F: **Relationships of active smoking to asthma and asthma severity in the EGEA study. Epidemiological study on the Genetics and Environment of Asthma.** *Eur Respir J* 2000, **15:**470–477.
10. Gilliland FD, Islam T, Berhane K, Gauderman WJ, McConnell R, Avol E, Peters JM: **Regular smoking and asthma incidence in adolescents.** *Am J Respir Crit Care Med* 2006, **174:**1094–1100.
11. Vignoud L, Pin I, Boudier A, Pison C, Nadif R, Le Moual N, Slama R, Makao MN, Kauffmann F, Siroux V: **Smoking and asthma: disentangling their mutual influences using a longitudinal approach.** *Respir Med* 2011, **105:**1805–1814.
12. Polonikov AV, Ivanov VP, Solodilova MA: **Genetic variation of genes for xenobiotic-metabolizing enzymes and risk of bronchial asthma: the importance of gene-gene and gene-environment interactions for disease susceptibility.** *J Hum Genet* 2009, **54:**440–449.

13. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25–29.
14. *The Gene Ontology database, version 1.8.* http://www.geneontology.org/, Date Accessed: 12/2012.
15. Islam T, Berhane K, McConnell R, Gauderman WJ, Avol E, Peters JM, Gilliland FD: **Glutathione-S-transferase (GST) P1, GSTM1, exercise, ozone and asthma incidence in school children.** *Thorax* 2009, **64:**197–202.
16. Islam T, McConnell R, Gauderman WJ, Avol E, Peters JM, Gilliland FD: **Ozone, oxidant defense genes, and risk of asthma during adolescence.** *Am J Respir Crit Care Med* 2008, **177:**388–395.
17. Castro-Giner F, Künzli N, Jacquemin B, Forsberg B, De Cid R, Sunyer J, Jarvis D, Briggs D, Vienneau D, Norback D, González JR, Guerra S, Janson C, Antó JM, Wjst M, Heinrich J, Estivill X, Kogevinas M: **Traffic-related air pollution, oxidative stress genes, and asthma (ECHRS).** *Environ Health Perspect* 2009, **117:**1919–1924.
18. Rogers AJ, Brasch-Andersen C, Ionita-Laza I, Murphy A, Sharma S, Klanderman BJ, Raby BA: **The Interaction of Glutathione S-transferase M1-null Variants with Tobacco Smoke Exposure and the Development of Childhood Asthma.** *Clin Exp Allergy* 2009, **39:**1721–1729.
19. Salam MT, Islam T, Gauderman WJ, Gilliland FD: **Roles of arginase variants, atopy, and ozone in childhood asthma.** *J Allergy Clin Immunol* 2009, **123:**1–8. 596–602, 602.
20. Wenten M, Gauderman WJ, Berhane K, Lin PC, Peters J, Gilliland FD: **Functional variants in the catalase and myeloperoxidase genes, ambient air pollution, and respiratory-related school absences: an example of epistasis in gene-environment interactions.** *Am J Epidemiol* 2009, **170:**1494–1501.
21. Breton CV, Salam MT, Vora H, Gauderman WJ, Gilliland FD: **Genetic variation in the glutathione synthesis pathway, air pollution, and children's lung function growth.** *Am J Respir Crit Care Med* 2011, **183:**243–248.
22. Elliott NA, Volkert MR: **Stress induction and mitochondrial localization of Oxr1 proteins in yeast and humans.** *Mol Cell Biol* 2004, **24:**3180–7.
23. Kaimul Ahsan M, Nakamura H, Tanito M, Yamada K, Utsumi H, Yodoi J: **Thioredoxin-1 suppresses lung injury and apoptosis induced by diesel exhaust particles (DEP) by scavenging reactive oxygen species and by inhibiting DEP-induced downregulation of Akt.** *Free Radic Biol Med* 2005, **39:**1549–1559.
24. Nickel C, Trujillo M, Rahlfs S, Deponte M, Radi R, Becker K: **Plasmodium falciparum 2-Cys peroxiredoxin reacts with plasmoredoxin and peroxynitrite.** *Biol Chem* 2005, **386:**1129–1136.
25. Tomita M, Okuyama T, Katsuyama H, Hidaka K, Otsuki T, Ishikawa T: **Gene expression in rat lungs during early response to paraquat-induced oxidative stress.** *Int J Mol Med* 2006, **17:**37–44.
26. Tseng CF, Huang HY, Yang YT, Mao SJT: **Purification of human haptoglobin 1–1, 2–1, and 2–2 using monoclonal antibody affinity chromatography.** *Protein Expr Purif* 2004, **33:**265–273.
27. IPA: *Ingenuity® Systems.* www.ingenuity.com, Date Accessed: 01/2013.
28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci USA* 2005, **102:**15545–15550.
29. *Molecular Signatures Database v3.1, updated Sep 27*; 2012. http://www.broadinstitute.org/gsea/, Date Accessed: 05/2013.

doi:10.1186/1476-069X-12-56
Cite this article as: Rava *et al.*: Selection of genes for gene-environment interaction studies: a candidate pathway-based strategy using asthma as an example. *Environmental Health* 2013 **12**:56.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central