Modelling of Genetic Interactions in GWAS Reveals More Complex Relations between Genotype and Phenotype

Joanna Zyla[1], Christophe Badie[2], Ghazi Alsbeih[3] and Joanna Polanska[1]
[1]*Data Mining Group, Faculty of Automatic Control, Electronic and Computer Science, Silesian University of Technology, Akademicka 16, Gliwice, Poland*
[2]*Cancer Genetics and Cytogenetics Group, Biological Effects Department, Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Didcot, OX11 ORQ, U.K.*
[3]*Radiation Biology Section, Biomedical Physics Dept., King Faisal Specialist Hospital & Research Centre, Riyadh 11211, Saudi Arabia*

Keywords: GWAS, Radiosensitivity, Polymorphism, Data Mining.

Abstract: The aim of this work is to present the complete methodology useful in GWAS analysis with small sample size, where comprehension of interaction between the genotype and phenotype is a main issue. By including all possible models of interaction into the process of model building, we were able to significantly increase the number of candidate polymorphisms and decrease the false discovery ratio.

1 INTRODUCTION

Every day our body is exposed to different types of radiation, which we can divided into two groups: ionizing radiation (IR) and non-ionizing one (nIR). Because of their high energy, all ionizing radiation cause the DNA damage in cells, while non-ionizing radiation doesn't damage DNA directly, affecting cells in other ways (Wrixon et al., 2004) (UNSCEAR, 2010). The ionizing radiation is one of the leading treatment of cancer though it causes many sides effects to the patient (Burnet et al., 2006) (NCI, 2012). People react to IR in different ways, some of them can stand very aggressive radiotherapy, while the others demonstrate high radio-intoxication just after the start of the treatment. This is why investigation of the individual reaction to ionizing radiation is so important to improve the overall health care level, to minimize the cost of the treatment and, what is the most important, to increase the patient comfort.

The reaction to IR is called radiosensitivity, which is defined as individual ability of cells, tissues, organs or organisms to the deal with the harmful effect of ionizing radiation. In 1906, Bergonie and Tribondeau find out that the radiosensitivity of cells is directly proportional to their activity and inversely proportional to the degree of differentiation (Bergoni and Tribondeau, 1906). Radiosensitivity has been shown to be heritable (Roberts et al., 1999). The investigation of biomarkers to asses the radiosensitivity is one of the leading problem in radiation biology. The most prominent biomarker is the activity of phosphorylation of histon H2AX, which gives a starting signal to double strand break repair (Taneja et al., 2004). However, it does not explain the observed diversity of radiosensitivity among patients, and there is still a big demand for search of new biomarkers. Main goal of this study is to demonstrate that proper techniques used to analyse the experimental data might lead to an extra discoveries by significant increase in the number of candidate biomarkers.

One of the way of polymorphism impact investigation is Genome-Wide Association Study (GWAS), which seems to be very efficient in search for genetic background of common diseases or phenomena like radiosensitivity. The first GWAS results were published in 2005 (Klein et al., 2005). The number of GWAS has been increasing since that, and, as it is stated in National Human Genome Research Institute database, for the day 25th of August 2013 there were 1716 GWAS-based publications, which report 11586 polymorphisms associated with diseases under investigation. The definition of GWAS says that the associated polymorphisms should be searched by genotyping at least 100,000 of them in the possible large population (Hindorff et al., 2009). In case of binary outcome, the most popular methods to detect and assess the risk alleles are the logistic regression and contin-

Zyla J., Badie C., Alsbeih G. and Polanska J..
Modelling of Genetic Interactions in GWAS Reveals More Complex Relations between Genotype and Phenotype.
DOI: 10.5220/0004807402040208
In *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms* (BIOINFORMATICS-2014), pages 204-208
ISBN: 978-989-758-012-3

gency tables (Bush and Moore, 2012)(Kamboh et al., 2012) (Chung et al., 2013). In case of quantitative trait tests like ANOVA are used. Because of the series of statistical test being performed, correction for multiple testing is necessary (Lin and Lee, 2012) (Pahl and Schafer, 2010).

In radiosensitivity the most important finding based on GWAS were published by Niu at el, where they show five genes (C13orf34, MALD2L1, PLK4, TPD52, and DEPDC1B) validated by siRNA knockdown experiment (Niu et al., 2010) and by Lin at el. which shows 12 significant polymorphisms related to the response to radiation by using the radiation hybryd network (Lin et al., 2005).

2 MATERIALS AND METHODS

2.1 Subject

The population under investigation is composed of 44 unrelated Caucasian individuals (uR).

2.2 Biological Experiments

Two types of information were collected per each participant. First one was the result of genotyping of 567,096 polymorphisms by Axiom myDesign arrays. Number of genotyped SNPs at each chromosome is presented in Table 1.

Table 1: Number of analysed polymorphisms by chromosomes.

Chr.	# of SNPs	Chr.	# of SNPs
1	41606	14	18302
2	45191	15	17030
3	39829	16	16376
4	37097	17	11655
5	35598	18	17749
6	43140	19	6899
7	30713	20	13297
8	31653	21	7894
9	26212	22	5523
10	27762	X	15790
11	25821	Y	2025
12	27262	MT	273
13	22397		
	TOTAL		567096

The second dataset includes activity of H2AX measured in two experimental conditions: 1) just after the irradiation of 2Gy and 2) in normal conditions. The irradiations were performed at room temperature with an A.G.O. HS X-ray system (Aldermaston, Reading, UK) (output 13 mA, 250 kV peak, 0.5 Gy/min for doses 0.5 4 Gy and 0.2 mA 4.9 mGy/min for doses up to 100 mGy). As the group o cells the T-lymphocyte cultures were prepared using the method described previously (O'Donovan et al., 1995; Finnon et al., 2008).

2.3 Models of Interaction

Three different models of SNP-gene expression interaction were investigated in this study. First of them is a genotype model, where each of the alleles gives independent level of gene expression (see Fig.1a). There is a special case of such association, called additive model, where each additional copy of the variant allele increases/decreases the response (Fig.1b). Another interaction model analysed in this study is a dominant one, where single variant allele impacts the gene expression level the same way as double variant allele form (Strachan and Read, 1999) (Fig.1c). The last model we distinguish is a recessive model, where only two variant alleles have impact on expression profile (Lewis, 2002) (Fig.1e).

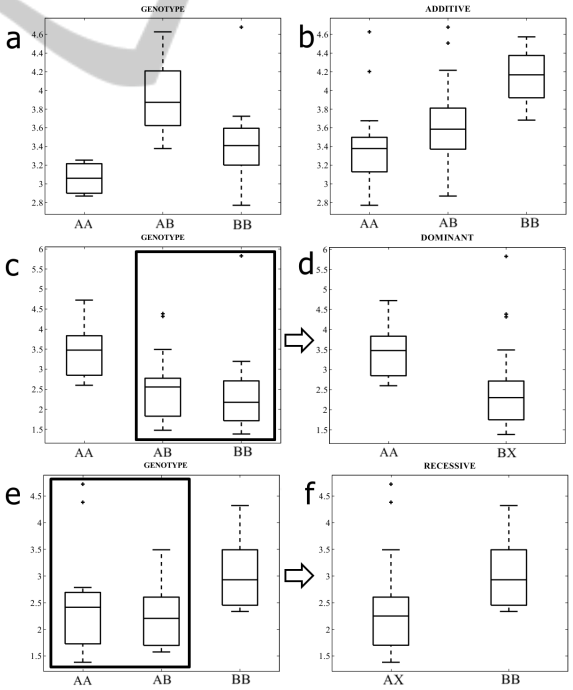

Figure 1: The boxplots of H2AX signal value presented according to the type of interaction model. Panel a) standard genotype interaction; b) additive interaction; c) boxplot of signal distibution in dominant interaction; d) final dominant model; e) boxplot of signal distibution in recessive interaction; f) final recessive model

2.4 Statistical Analysis

Since the measurements of H2AX phosphorylation were taken in two experimental conditions, the normalized values of signal induction after IR were calculated with the use of reference genes and all of the analyses were performed at two endpoints: 0Gy and logarithm of fold change (FCH). The normality of signal distribution was verified by Lilliefors test. Homogeneity of variance was checked by Bartlett's test for genotype model, and F test for dominant and recessive models. The parametric tests for equivalence of population mean values were used (ANOVA for genotype, t-test for dominant and recessive models). In the next step, the best interaction model per each polymorphism at every endpoint was chosen with the use of minimum p-value criterion. Per every SNP, the minimum p-value was chosen from the set of three p-values obtained for the analysed models of interaction: genotype (p_G), dominant (p_D), and recessive (p_R) - Figure 2.

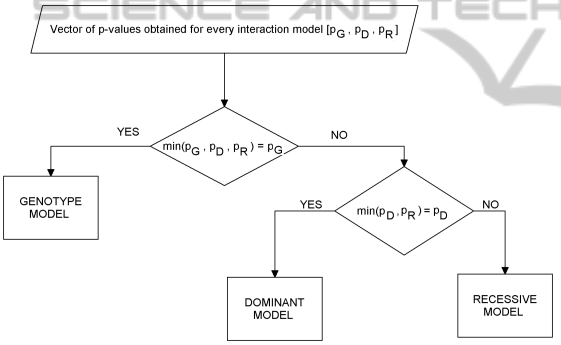

Figure 2: Algorithm of decision making on the final model of interaction for particular SNP at chosen endpoint with the use of minimum p-value critertion.

As a set of candidate polymorphisms those with minimum p-value$<$0.05 were considered. The false discovery ratio was estimated to accompany set of candidate polymorphisms.

3 RESULTS AND DISCUSSION

Table 2 presents the summary of genotyped data. The results of normality testing presents table 3.

One can observe that the number of null hypothesis rejections is lower than the expected by chance, which leads to the general acceptance of null hypothesis on normality of signal distribution. While looking at the results of the verification of hypothesis on homogeneity of variances (see table 4), the observed number of hypothesis rejections exceeds the expected by chance, which forces the use of proper parametric tests with correction for unequality of variances.

Finally, the hypothesis on population mean value equality was checked, separate analyses were performed for all three models of interactions. The summary of these is presented in table 5.

The criterion of minimal p-value was used to get the final model for particular SNP-endpoint interaction, the results are presented in table 6. In tested his-

Table 2: Summary of polymorphism genotyping in group under investigation.

Genotyping results	# of SNPs	%
single form only	33,679	5.93%
two different forms	235,007	41.44%
all three forms	147,413	25.99%

Table 3: Results of testing on normality of signal distribution.

		Model of interaction		
	# of tests	Genotype 442,245	Dominant 558,028	Recessive 636,418
0Gy p$<$0.05	N	13,502	11,262	12,686
	%	3.05	2.02	1.99
	FDR	100	100	100
FCH p$<$0.05	N	19,545	17,637	20,274
	%	4.42	3.16	3.19
	FDR	100	100	100

Table 4: Results of testing on homogeneity of variances.

		Model of interaction		
	Total	Genotype 147,415	Dominant 279,014	Recessive 318,209
0Gy p$<$0.05	N	5,354	11,130	12,837
	%	3.63	3.98	4.03
	FDR	100	100	100
FCH p$<$0.05	N	8,107	16,199	18,537
	%	5.5	5.81	5.83
	FDR	100	86.12	85.83

Table 5: The results of comparison study of H2AX at both endpoints - 0Gy and FCH.

		Model of interaction		
	Total	Genotype 147,415	Dominant 279,014	Recessive 318,209
0Gy p$<$0.05	N	6,947	26,668	30,641
	%	4.71	9.56	9.63
	FDR	100.00	52.31	51.93
FCH p$<$0.05	N	7,177	27,675	31,579
	%	4.86	9.92	9.92
	FDR	100.00	50.41	50.38

ton phosphorylation activity majority of SNPs represent dominant or recessive model, and very few (with high FDR) represent genotype interaction.

Table 6: The results of the final optimal model selection for H2AX at both endpoints - 0Gy and FCH.

		Optimal model of interaction		
		Genotype	Dominant	Recessive
0Gy	N	1,159	25,052	29,078
	FDR	100	55.69	54.72
FCH	Nl	1,050	26,040	29,901
	FDR	100	53.57	53.21

All the polymorphisms with minimal p-value for final model being less than 0.05 were considered as the candidate SNPs related to the radiosensitivity phenomena. If genotype model was only considered, there would be 1159 candidate SNPs for 0Gy and 1050 candidate SNPs for FCH found with FDR equal to 100%. If additional models of interaction, recessive and dominant, are included into the study together with the proposed method for model selection, the number of candidate SNPs increases significantly to 55,289 giving FDR at the level of 67.34%. Only 2.1% of final models are genotype interactions at 0Gy and 1.84% for FCH endpoint, while 45.31% and 45.69% respectively are of dominant type. The most frequent for both endpoints is recessive model with 52.59% at 0Gy and 52.46% for FCH of candidate SNPs.

The filtration of candidate SNPs could be done in several ways, some might apply the standard statistical corrections for multiple testing, another might go for candidate SNP validation done with independent sample, the others might validate SNPs by the functional analysis. Due to the limited power of statistical tests, resulting from relatively small sample size, the application of multiple testing correction techniques is not recommended. We do propose to check on their genome location and functional class. Since the nonsynonymous SNPs (nSNP), are the most interesting because they directly affect translated amino-acid protein sequence (Ramensky et al., 2002), the SNP validation might be done by requiring from candidate SNP to be nSNP. The average number of nonsynonymous SNPs is 1.32% and 1.23% (for 0Gy and FCH endpoints respectively), and the average percentage of polymorphisms in functional regions (intron, exon, nsSNP, UTR etc.) is equal to 41.36% and 41.27% respectively, table 7.

Table 7: Functional description of obtained candidate SNPs.

	# SNPs	# Functional SNPs	(%)
0Gy	55,289	22,868	41.36
FCH	56,991	23,519	41.27
intersection	7,956	3,253	14.20

4 CONCLUSIONS

The analysis revealed that different models of interaction must be included in the investigation of the genetic background of biological phenomena, especially in the case of studies with limited sample size. Additionaly, looking at allelic frequencies and genotyping results only would limit the findings of the best candidate biomarkers. The presented study demonstrated that it is possible to design the proper statistical analysis strategy even for small sample size decreasing significantly the false discovery rate for the set of candidate SNPs. Due to limited power of statistical tests applied, validation of candidate SNPs must be performed by functional analysis and/or independent validation experiment.

ACKNOWLEDGEMENTS

We would like to thank Dr. S. Majid, Ms. N. Al-Harbi, Ms. S. Al-Qahtani for running the Axiom Affymetrix platform and Sylwia Kabacik and Paul Finnon for run the H2AX test. The work was financially supported by NCN grant HARMONIA UMO-2013/08/M/ST6/00924 (JP), the National Institute for Health Research Centre for Research in Public Health Protection at Public Health England (CB), the National Science, Technology & Innovation Plan (NSTIP) Project 11-BIO1429-20 (KFSHRC RAC# 2120 003) (GA), and SUT grant BK/219/RAU-1/2013/t.10 (JZ). Additionally, Joanna Zyla is holder of scholarship DoktoRis - Scholarship program for Innovative Silesia.

REFERENCES

Bergoni, J. and Tribondeau, L. (1906). De quelques rsultats de la radiotherapie et essai de fixation d'une technique rationnelle. *Comptes-Rendus des Sances de l'Acadmie des Sciences*, 43:983–985.

Burnet, N., Elliott, R., Dunning, A., and West, C. (2006). Radiosensitivity, radiogenomics and rapper. *Clinical Oncology*, 18(7):525–528.

Bush, W. and Moore, J. (2012). Chapter 11: Genome-wide association studies. *PLOS Comput Biol.*, 8(12).

Chung, S., Low, S., Zembutsu, H., Takahashi, A., ans M. Sasa, M. K., and Nakamura, Y. (2013). A genome-wide association study of chemotherapy-induced alopecia in breast cancer patients. *Breast Cancer Res.*, 15(5):R81.

Finnon, P., Robertson, N., Dziwura, D., Raffy, C., Zhang, W., Ainsbury, L., Kaprio, J., Badie, C., and Bouffler, S. (2008). Evidence for significant heritability of apoptotic and cell cycle responses to ionising radiation. *Hum Genet.*, 123(5):485–493.

Hindorff, L., Sethupathy, P., Junkins, H., Ramos, E., Mehta, J., Collins, F., and Manolio, T. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS*, 106(23):9362–9367.

Kamboh, M., Demirci, F., Wang, X., Minster, R., Carrasquillo, M., Pankratz, V., Younkin, S., Saykin, A., Jun, G., Baldwin, C., Logue, M., Buros, J., Farrer, L., Pericak-Vance, M., Haines, J., Sweet, R., Ganguli, M., Feingold, E., Dekosky, S., Lopez, O., and Barmada, M. (2012). Genome-wide association study of alzheimer's disease. *Transl Psychiatry.*, 15:2:e117.

Klein, R., Zeiss, C., Chew, E., Tsai, J., Sackler, R., Haynes, C., Henning, A., SanGiovanni, J., Mane, S., Mayne, S., Bracken, M., Ferris, F., Ott, J., Barnstable, C., and Hoh, J. (2005). Complement factor h polymorphism in age-related macular degeneration. *Science*, 308(5720):385–389.

Lewis, C. (2002). Genetic association studies: design, analysis and interpretation. *Brief Bioinform.*, 3(2):146–153.

Lin, A., Wang, R., Ahn, S., Park, C., and Smith, D. (2005). A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res*, 20(8):1122–1132.

Lin, W. and Lee, W. (2012). Improving power of genome-wide association studies with weighted false discovery rate control and prioritized subset analysis. *PLOS One.*, 7(4).

NCI (2012). *Radiation Therapy and You*. National Institute of Health publiaction No. 12-7157.

Niu, N., Qin, Y., Fridley, B., Hou, J., Kalari, K., Zhu, M., Wu, T., Jenkins, G., Batzler, A., and Wang, L. (2010). Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome Res*, 20(11):1482–1492.

O'Donovan, M., Freemantle, M., Hull, G., Bell, D., Arlett, C., and Cole, J. (1995). Extended-term cultures of human t-lymphocytes: a practical alternative to primary human lymphocytes for use in genotoxicity testing. *Mutagenesis.*, 10(3):189–201.

Pahl, R. and Schafer, H. (2010). Permory: an ld-exploiting permutation test algorithm for powerful genome-wide association testing. *Bioinformatics*, 26(17):2093–2100.

Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human nonsynonymous snps: server and survey. *Nucleic Acids Research*, 30(17):3894–3900.

Roberts, S., Spreadborough, A., Bulman, B., Barber, J., Evans, D., and Scott, D. (1999). Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer? *American Journal of Human Genetics*, 65(3):784–794.

Strachan, T. and Read, A. (1999). *Human Molecular Genetics*. Wiley-Liss, New York, 2nd edition.

Taneja, N., Davis, M., Choy, J., Beckett, M., Singh, R., Kron, S., and Weichselbaum, R. (2004). Histone h2ax phosphorylation as a predictor of radiosensitivity and target for radiotherapy. *J Biol Chem.*, 279(3):2273–2280.

UNSCEAR (2010). *Sources and effects of ionizing radiation*. United Nations Publication, New York.

Wrixon, A., Barraclough, I., Clark, M., Ford, J., Diesner-Kuepfer, A., and Blann, B. (2004). *Radiation, People and the Enviroment*. International Atomic Energy Agency, Austria.