Pharmacogenomics and Personalized Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

SHORT REPORT

Lack of correlation between in silico projection of function and quantitative real-time PCR-determined gene expression levels in colon tissue

This article was published in the following Dove Press journal:
Pharmacogenomics and Personalized Medicine
6 September 2013
Number of times this article has been viewed

Rosalind B Penney[1]
Abbie Lundgreen[2]
Aiwei Yao-Borengasser[3]
Vineetha Koroth-Edavana[3]
Suzanne Williams[3]
Roger Wolff[2]
Martha L Slattery[2]
Susan Kadlubar[3]

[1]Department of Environmental and Occupational Health, University of Arkansas for Medical Sciences, Little Rock, AR, [2]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT, [3]Division of Medical Genetics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Abstract: There are a number of in silico programs that use algorithms and external web sources to predict the effect of single nucleotide polymorphisms (SNPs). While many of these programs have been shown to predict accurately the effect of SNPs in functional areas of the gene, such as 5′ upstream or coding regions, empiric research may be warranted to confirm the functional consequences of SNPs that are predicted to have little to no effect. We compared predictions from FASTSNP (Function Analysis and Selection Tool for Single Nucleotide Polymorphism) and F-SNP (Functional Single Nucleotide Polymorphism) with experimentally derived genotype-phenotype correlations to determine the accuracy of these programs in predicting SNP functionality. We used normal colon tissue to evaluate 24 TagSNPs within six genes. Two of 16 SNPs that were predicted to have no functional effect in FASTSNP were significantly associated with gene expression. Only one of the eight SNPs that were predicted to have a low to high effect was significantly associated with gene expression. While the two in silico programs that were used were similar in their results for the SNPs predicted by FASTSNP to have no effect, of SNPs with scores from low to high, there were three that received an F-SNP score below what is considered functionally significant. In silico programs can fail to identify functional SNPs, supporting a continuing role for empiric analysis of SNP function. Laboratory analysis is necessary to identify causal SNPs accurately, establish biological plausibility of the effect, and ultimately inform cancer prevention strategies.

Keywords: in silico prediction, colon, single nucleotide polymorphisms

Introduction

The ability to link functional genetic variants with disease risk leads to advances in diagnostics and therapeutics.[1] Over 10 million single nucleotide polymorphisms (SNPs) have been reported,[2] with an estimated 100,000–300,000 that alter an amino acid.[3] In silico prediction programs have been developed to identify SNPs with possible functional effects.[4] Several programs are available, each with unique algorithms to assess the potential effect of an amino acid sequence substitution.[5] For instance, FASTSNP (Function Analysis and Selection Tool for Single Nucleotide Polymorphism) utilizes web wrapper agents to gather information from 11 different web servers to offer real-time information on phenotypic risk and functional effects, and F-SNP (Functional Single Nucleotide Polymorphism) uses 16 different tools and databases in an integrated fashion to predict functionality based on splicing, transcription, translation, and post-translation.[4,6]

These programs are useful in prioritizing SNPs for genotyping, as well as for more detailed functional analyses. A large survey of many of these programs showed

Correspondence: Susan Kadlubar
University of Arkansas for Medical Sciences, 4301 W Markham, 580 Little Rock, AR 72205, USA
Tel +1 501 526 7957
Email sakadlubar@uams.edu

a high level of consistency between programs in identifying high-risk/high-priority SNPs for colon cancer research.[7] However, evolving research supports a functional role for intronic SNPs. For example, an intronic SNP associated with acute lung injury and asthma regulates promoter activity of smMLCK,[8] another in *PRRX2* has been shown to interact with the conditioning region in KLK2-KLK3,[9] and yet another in the *GHI* gene that is associated with reduced colorectal cancer risk was shown to decrease GHI expression.[10] Each of these intronic SNPs is predicted to have no to low risk of effect in either the in silico FASTSNP or F-SNP prediction programs.

To explore the accuracy of predictive models with SNP functionality, identified tagSNPs were correlated with gene expression in normal colon tissue. Empiric results were then compared with the in silico risk prediction programs, FASTSNP and F-SNP.

Materials and methods

Tissue samples

Deidentified normal frozen colon tissues (n = 82) were obtained from the Cooperative Human Tissue Network, funded by the National Cancer Institute, and stored at −80°C. Of the sample population, 54% were male and 46% were female. The tissue donors were aged 17–92 (mean 60.48) years and were of Caucasian (n = 51), African American (n = 23), Asian (n = 1), and unknown (n = 7) origin.

Reverse transcription and quantitative real-time polymerase chain reaction

Total DNA was isolated from normal colon tissue samples using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Valencia, CA, USA). Total RNA was isolated utilizing Trizol (Invitrogen, Grand Island, NY, USA) for homogenization, and the RNEasy Mini kit (Qiagen) for isolation using a protocol developed by Mauricio Rodriquez-Lanetty (unpublished) with minor alterations. Briefly, tissues (about 25 mg) were homogenized in 150 µL Trizol using a Bullet Blender and stainless steel beads. The homogenate was placed in a new vial with 450 µL of Trizol. After adding 100 µL of chloroform, the vials were shaken well, incubated for 2 minutes at room temperature, centrifuged, and the supernatant was placed in a new vial. Equal parts of 100% ethanol were added, and the mixture placed in an RNEasy spin column. RNA was washed and eluted according to the RNEasy protocol.

First strand cDNA synthesis was performed using the High Capacity RNA-to-cDNA kit (ABI, Carlsbad, CA, USA) on 500 ng total RNA, as measured by an RNA 6000 Nano kit (Agilent, Santa Clara, CA, USA). Quantitative real-time reverse transcription polymerase chain reaction (PCR) reactions were performed on the ABI 7900HT Fast Real Time PCR System using Taqman primer/probe sets and Taqman Fast Universal PCR Master Mix no AmpErase® UNG (ABI). Experiments were run as per the manufacturer's protocol in triplicate on cDNA diluted 1:10 for 50 PCR cycles, retaining those with standard deviations <1 (exclusions: *IFNGR2* [1], *IL1B* [1]). Samples were normalized to *β-actin*, discarding those with *β-actin* Ct (cycle threshold) >30 (*IFNGR1* [4], *IFINGR2* [4], *IL1B* [5], *LEPR* [1], *RPS6 KB1* [1], *TSC2* [4]). Genes of interest Ct ≥40 or undetermined were set to 40. *β-actin* was chosen as the housekeeping gene because, in normal colon tissue, it has been shown that structural housekeeping genes such as *β-actin* have less variation than metabolic housekeeping genes such as glyceraldehyde 3-phosphate dehydrogenase.[11]

TagSNP selection and genotyping

TagSNPs were selected using the following parameters: r^2 = 0.8 defined LD blocks using a Caucasian LD map, minor allele frequency >0.1, range −1,500 base pairs from the initiation codon to +1,500 base pairs from the termination codon, and one SNP/LD bin. All markers were genotyped using a multiplexed bead array assay format based on GoldenGate chemistry (Illumina, San Diego, CA, USA). A genotyping call rate of 99.93% was attained. Blinded internal replicates represented 1.6% of the sample set. The duplicate concordance rate was 99.996%.

In silico prediction programs

Two in silico programs were used. FASTSNP is a web-based tool for assessing phenotypic effects of SNPs through the use of external web servers and a prediction algorithm. FASTSNP uses a ranking system from 0 (no known effect) to 5 (very high risk) based on location of the SNP (eg, 5′ upstream, 3′ untranslated region, intronic) and possible functional effects such as amino acid changes, alterations in splicing sites, and "premature translation termination".[6] F-SNP also utilizes bioinformatic tools and websites to predict the functional effects of SNPs. The process has several steps, with each step determining the next. For instance, if a mutation is found in the coding region through Ensembl, the information is then submitted to an outside bioinformatics website, such as PolyPhen, to test for functional effect.[4]

Statistical analysis

Identified TagSNPs for 34 genes (*CYP19A1, IFNG, IFNGR1, IFNGR2, IKBKB, IL10, IL15, IL17A, IL1A, IL1B, IL1RN*, *IL2*,

IL23R, IL2RA, IL4, IL6, IL6R, IL8, LEPR, MTOR, NFKB1, PDGFB, PDK1, PIK3CA, PRKAG2, PTEN, RPS6KB1, RPS6KB2, STAT3, STAT 5B, TGFB1, TNF, TSC2, VEGFA) were entered into the FASTSNP website, and predicted risk values were noted. Six genes (*IFNGR1, IFNGR2, IL1B, LEPR, RPS6KB1*, and *TSC2*) were identified as having SNPs that were predicted to have a score of either 2–3 or 3–4 (low to medium or medium to high risk of effect, respectively). From these six genes, tagSNPs with a score of 0–0 (no or unknown risk, n=16) or with a score of either 2–3 or 3–4 (low to high risk, n=8) were chosen for further comparison with phenotype data. Results from F-SNP were based on transcriptional regulation and marked either "changed"/"not changed" or "exist"/"not exist." A functional significance score is given, with a score of ≥0.5 being considered likely to lead to functional changes.[12] The TagSNPs chosen for FASTSNP prediction were entered into the F-SNP prediction program and compared with both phenotype data and with FASTSNP predictions in order to assess similarity between prediction programs.

Statistical analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC, USA). The level of expression for the candidate gene was calibrated to the expression of the housekeeping gene to generate change in Ct. Expression levels were calculated by taking 2^ΔCt and the median of those values was assessed by genotype. A codominant model was initially assumed, but if a dominant or recessive model fitted the data better, that model was evaluated and is presented. *P*-values comparing median expression levels across genotypes are based on Wilcoxon rank-sum and Kruskal–Wallis rank-sum tests. Statistical significance was set at $P < 0.05$. SNP associations were performed among Caucasians and African Americans separately, and the directions of the associations are the same for both races for the three leptin receptor SNPs that were reported as being significant (rs8179183, rs9436301, rs4655537). Race was not associated with gene expression. Expression was also not statistically significantly different by age or gender.

Results

Predicted and actual effects in normal colon samples

The predicted FASTSNP and F-SNP effects and gene expression association *P*-values of the 24 TagSNPs are presented in Table 1. Of 16 SNPs predicted to have no/unknown (0–0) effect, two (*LEPR* rs4655537 and rs9436301) were found to be significantly associated with gene expression (Table 2). The common homozygous *LEPR* rs4655537 genotype (GG) is associated with a 1.7-fold increase ($P = 0.01$) in expression of LEPR compared with the heterozygous or homozygous variant (GA/AA) genotype. The CC variant *LEPR* rs9436301 genotype is associated with a 1.52-fold increase in gene expression ($P = 0.04$) as compared with the CT/TT genotype.

Table 1 Prediction scores and association with gene expression

Gene	SNP	FASTSNP score	F-SNP score	*P*-value for SNP association with expression
IFNGR1	rs1327475	2-3	0.176	0.26
	rs9376267	0-0	0.208	0.90
IFNGR2	rs9808753	3-4	0.633	0.35
	rs9976971	0-0	0.5	0.52
IL1B	rs1143634	2-3	0.330	0.92
	rs1143633	0-0	0.268	0.29
LEPR	rs1137101	3-4	0.291	0.28
	rs8179183	3-4	0.533	0.048
	rs1805096	2-3	0.5	0.15
	rs12145690	0-0	0.217	0.83
	rs9436301	0-0	0.141	0.04
	rs6704167	0-0	0.176	0.87
	rs1171271	0-0	0.242	0.84
	rs6673324	0-0	0.109	0.78
	rs12059300	0-0	0.065	0.83
	rs4655537	0-0	0.158	0.01
	rs1938484	0-0	0.242	0.28
RPS6KB1	rs180523	3-4	1	0.63
	rs8071475	0-0	0.208	0.20
	rs180515	0-0	0.276	0.42
TSC2	rs1051771	2-3	0.568	0.99
	rs2073636	0-0	0.242	0.74
	rs30259	0-0	0.176	0.12
	rs3087631	0-0	0.050	0.19

Abbreviations: SNP, Single Nucleotide Polymorphism; FASTSNP, Function Analysis and Selection Tool for Single Nucleotide Polymorphism; F-SNP, Functional Single Nucleotide Polymorphism.

Of the eight tagSNPs that were predicted to have a low to high effect (2–3 or 3–4) in the FASTSNP program, only *LEPR* rs8179183 was significantly associated with gene expression. The common homozygous genotype (GG) was associated with a 1.6-fold decrease ($P = 0.048$) in expression compared with the heterozygote and homozygous variant (GC/CC).

When compared, FASTSNP and F-SNP scores were similar, although not entirely consistent (Table 1). For TagSNPs that were predicted to have no (0–0) effect in FASTSNP, the F-SNP score was below 0.5, the score at which a SNP is likely to lead to functional changes. Of the eight SNPs that were predicted to have a low to medium (2–3) or medium to high (3–4) effect with FASTSNP, five received a functional significance score ≥0.5. The other three ranged in scores from 0.176 to 0.330, causing their prediction to match the genotype/phenotype results better. While four of the five

Table 2 SNPs with significant association with gene expression

Gene	SNP	N	Gene expression[a]	Kruskal–Wallis *P*-value	FASTSNP score	F-SNP score
LEPR	rs8179183					
	GG	54	40.6433	0.048	3-4	0.533
	GC/CC	27	65.1387			
	rs9436301					
	TT/TC	70	44.3307	0.043	0-0	0.141
	CC	11	67.5232			
	rs4655537					
	GG	36	59.7931	0.011	0-0	0.158
	GA/AA	45	33.6846			

Note: [a]Gene expression values are median $2^{\wedge}\Delta Ct \times 10^4$.

Abbreviations: SNP, Single Nucleotide Polymorphism; FASTSNP, Function Analysis and Selection Tool for Single Nucleotide Polymorphism.

tagSNPs with a functional significance score ≥0.5 hovered near 0.5 (0.5–0.633), one (*RPS6KB1* rs180523) had a functional significance score of 1. *RPS6KB1* rs180523 also had a FASTSNP score of 3–4, but the expression results showed no statistically significant differences in expression across genotypes ($P = 0.63$).

Discussion

Differentiating between SNPs that may be deleterious and those that are "benign" is critical to risk assessment and the design of cancer prevention strategies.[5] With the human genome being home to potentially millions of SNPs, laboratory discovery of individual SNPs is a daunting task. For this reason, in silico programs have emerged to assist in choosing functional SNPs. These programs use readily available scientific data and bioinformatics to offer predictions on the functional effects of SNPs. This study sought to determine genotype-phenotype relationships empirically, and found that a zero risk of effect in an in silico prediction program does not guarantee a lack of effect of certain SNPs in human colon samples.

In an effort to explore this in relation to gene expression, 82 colon samples were genotyped and phenotyped for the 24 TagSNPs predicted by FASTSNP to have either no effect (0–0) or a low to medium or medium to high effect (2–3 or 3–4, respectively). Our results showed that two of the 16 SNPs that were predicted to have no effect had a significant association with gene expression. In the eight SNPs with a predicted low to high effect, only one showed a significant association with gene expression.

Not all prediction programs generate similar results. The databases and external websites employed by each program are different (although there is some overlap), and unique algorithms are likely to generate disparate results. Thus, FASTSNP results were compared with those of F-SNP. F-SNP combines accumulated results into a single "functional significance score," with a score of ≥0.5 considered likely to lead to functional changes, given that that is the median score for known disease-related SNPs.[4] For these data, FASTSNP and F-SNP scores corresponded for SNPs predicted to have no known effect. However, they did not match with all SNPs that were predicted to have a low to high effect.

There is a chance that the lack of correlation is due to the small sample size. Also, the functionality of SNPs is not limited to RNA expression, and prediction programs are designed to explore other dimensions of functionality, such as amino acid changes and alterations in splicing sites. This may explain a portion of the high-priority SNPs that showed no change in mRNA expression. Further functionality experiments would be necessary to explore other mechanisms of action, such as post-translational modification, protein expression, and protein function, specifically with the leptin receptor protein. There may also be organ-specific differences in gene expression, which may have impacted the results shown here. This further necessitates laboratory functionality studies and inspection of low-priority SNPs in a case-by-case manner. It is also possible that the SNPs chosen for analysis are not truly functional SNPs, but exist in tight linkage with the causative SNP. For this reason also, biochemical studies are necessary to define the mechanistic basis of the noted associations.

There are a few examples of comparison of FASTSNP and functional in vitro experiments. However, these only focus on the high-priority SNPs. For example, a study in the Chinese Han population found two cystathionine gamma-lyase SNPs (rs482843 and rs1021737) to be identified by FASTSNP as high-priority SNPs, yet which showed no significant contribution to the risk of essential hypertension in this population.[13] On the other hand, one in vitro study created a p16INK4A protein (from the *CDKN2A* gene) based on SNPs identified as high-priority by FASTSNP and other in silico programs, and found that *CDKN2A* rs11552822 may

lead to a decrease in binding affinity for CDK6, and may be involved in the development of malignant melanoma.[14]

In silico programs have been shown to be accurate when predicting functional effects with SNPs that rank very high on their prediction list, and certainly these higher-risk SNPs may be prioritized in laboratory-based research. However, it is not likely that they stand alone in the progression of complex disease.[15] Thus, SNPs that are ranked as "no risk" by in silico programs may actually have an effect on gene expression, which may, in turn, lead to an effect on protein abundance and subsequent functioning of the enzyme. For example, the no to low priority *GH1* rs2665802 has been associated with both a decrease in human growth hormone gene expression and growth hormone secretion. It was noted that this SNP may work in conjunction with other SNPs not studied, but the contribution of the SNP was found to be direct.[10]

Even low to medium effects on enzymatic activity may play an important role in the development of disease. Therefore, functional analyses of these low risk SNPs are necessary to capture fully the genotypic contributions to phenotype. This information is critical in determining the biological basis of variability, and can potentially aid in the design of rational intervention/prevention strategies.

Acknowledgment

This work was supported by R01 CA48998 (MLS).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Andersen MC, Engstrom PG, Lithwick S, et al. In silico detection of sequence variations modifying transcriptional regulation. *PLoS Comp Biol*. 2008;4(1):e5.
2. Lee JE. High-throughput genotyping. *Forum Nutr*. 2007;60:97–101.
3. Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR. Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genet*. 2005;1(6):e83.
4. Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res*. 2008;36:D820–D824.
5. Wang LL, Li Y, Zhou SF. A bioinformatics approach for the phenotype prediction of nonsynonymous single nucleotide polymorphisms in human cytochromes P450. *Drug Metab Dispos*. 2009;37(5):977–991.
6. Yuan HY, Chiou JJ, Tseng WH, et al. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res*. 2006;34:W635–W641.
7. George Priya Doss C, Rajasekaran R, Arjun P, Sethumadhavan R. Prioritization of candidate SNPs in colon cancer using bioinformatics tools: an alternative approach for a cancer biologist. *Interdiscip Sci*. 2010;2(4):320–346.
8. Han YJ, Ma SF, Wade MS, Flores C, Garcia JG. An intronic MYLK variant associated with inflammatory lung disease regulates promoter activity of the smooth muscle myosin light chain kinase isoform. *J Mol Med (Berl)*. 2012;90(3):299–308.
9. Ciampa J, Yeager M, Amundadottir L, et al. Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study. *Cancer Res*. 2011;71(9):3287–3295.
10. Millar DS, Horan M, Chuzhanova NA, Cooper DN. Characterisation of a functional intronic polymorphism in the human growth hormone (GH1) gene. *Hum Genomics*. 2010;4(5):289–301.
11. Rubie C, Kempf K, Hans J, et al. Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes*. 2005;19(2):101–109.
12. Lee PH, Shatkay H. An integrative scoring system for ranking SNPs by their potential deleterious effects. *Bioinformatics*. 2009;25(8):1048–1055.
13. Li Y, Zhao Q, Liu XL, et al. Relationship between cystathionine gamma-lyase gene polymorphism and essential hypertension in Northern Chinese Han population. *Chin Med J (Engl)*. 2008;121(8):716–720.
14. Rajasekaran R, Priya Doss CG, Sudandiradoss C, Ramanathan K, Sethumadhavan R. In silico analysis of structural and functional consequences in p16INK4A by deleterious nsSNPs associated CDKN2A gene in malignant melanoma. *Biochimie*. 2008;90(10):1523–1529.
15. Prokunina L, Alarcon-Riquelme ME. Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Rev Mol Med*. 2004;6(10):1–15.