Doctoral School of Health Sciences

Faculty of Health Sciences, University of Pécs

Head of Doctoral School: Prof Dr. Kiss, István MD, Ph.D., DSc.

RESEARCH ON THE APPLICATION OF MOLECULAR EPIDEMIOLOGY IN THE PREDICTION AND PROGNOSIS OF CANCER

Doctoral (Ph.D.) thesis booklet

Lu Zhang

Programme Leader: Prof. Dr. Kiss, István

Supervisor: Prof. Dr. Kiss, István

Pécs, 2024

1 Introduction

Molecular epidemiology is a subject that integrates molecular biology and basic epidemiology. It introduces biomarkers that are involved in the molecular pathway and specific genes related to disease risk into epidemiology to recognize disease causation, risk factors, prevention, and even treatment. It emphasizes the interactions between genetic, environmental and other factors that result in diseases. Cancer is a group of diseases caused by abnormal cells creating rapidly and uncontrollably and spreading to other organs. It is the world's second leading cause of death. Cancer is caused by a combination of genetic and environmental factors.

Molecular epidemiology provides tools for understanding the interaction between these factors. Different from the "traditional epidemiology" that revolves around time, place, and person, molecular epidemiology pays attention to looking for biomarkers related to diseases, such as DNA, transcription factors, RNA, cell surface receptors, enzymes, and even metabolites, and then uses them to explain the diseases' mechanisms in populations. According to the stages from exposure to cancer development, biomarkers used in cancer molecular epidemiology research can be categorized into markers of exposure, internal dose markers, biologically effective dose markers, early biological effect markers, altered structure/function markers, biomarkers of prognosis, and markers of disease. Those markers play an important and essential role in the diseases' early detection, diagnosis, staging, treatment, and prognosis.

However, as a novel field, how molecular epidemiology is used in specific practical studies, and how it helps to discover causes, explore risk factors, and ultimately protect, still remains an unclear explanation. Hence, the aim of this study is to display how molecular epidemiological runs in cancer research through two practical studies: ‘The relationship between single nucleotide polymorphisms and skin cancer susceptibility’ and ‘The treatment effect of hydrogen gas on lung cancer’.

2 The therapeutic effect of hydrogen gas on lung cancer -- Linc-PINT and LincRNA-p21 as biomarkers

2.1 Background

Lung cancer (LC), as one of the most common cancer, causes increased numbers of morbidity and mortality all over the world. The long intergenic non-protein coding RNA p53-induced transcript (LINC-PINT) and LincRNA-p21 are TP53-induced transcripts. It was investigated that the expression of LINC-PINT and lincRNA-P21 decreased in the malignant cancer. Molecular hydrogen (H_2) is a new medical gas that is used as a selective antioxidant in the anti-inflammation and anti-apoptosis functions modulating. H_2 also plays a role in the LC treatment. Therefore, the aim of this study is to explore the treatment effect of H_2 in lung cancer cells by identifying the expression of LINC-PINT and lincRNA-P21.

2.2 Method

Electrochemical water device was utilizted to produce hydrogen gas. qRT-PCR was employed to assess the expression of LINC-PINT and lincRNA-P21 in lung cancer cells, respectively. One-way analysis of variance (ANOVA) and linear regression were carried out for analysing multiple groups' differences and associations.

2.3 Results

2.3.1 Investigation of possible influencing factors

To ensure that flow of H_2 gas was stable, we investigated the relationship between H_2 and time. The volume of the produced hydrogen had a linear relationship with time ($Y=32.78+1.55X$, $R^2=97.60\%$, $P<0.05$). Besides, ANOVA result showed that there was no statistical difference between the H_2 concentration of the three layers of the box ($F=0.589$, $P>0.05$). Thus, we confirmed, that the middle and the top layers of the culture

dish were exposed to equal concentrations of hydrogen gas.

2.3.2 Effect of H_2 concentration on lincRNAs expression

Lung cancer cells were treated with 3 different concentrations (0%,5%,and 10%) of H_2 gasfor three different time periods (2H, 3H, and 3H30). The expression levels of LINC-PINT were significantly correlated with the increasing H_2 concentrations, from 0% (control group) to the 5% and 10%-concentrations in the 2H30 and 3H time groups (R^2_{2H30}=0.52, R^2_{3H}=0.57, both P<0.05). The positive relationship between expression levels and H_2 concentrations was also observed for lincRNA-P21 expression in the 2H30 time group (R^2=0.88, P<0.01). In contrast, the expression of LINC-PINT showed a significant negative correlation with H_2 concentration in the 3H40 time group (R^2=0.81, P<0.01).

2.3.3 Effect of treatment time on linc-RNA expression

The expression levels of LINC-PINT in the 5% H_2 and the 10% H_2 group followed opposite trends after different lengths of treatment, however neither trend was statistically significant ($F_{5\%H2}$=1.60 and $R^2_{5\%H2}$=0.78, $F_{10\%H2}$=3.64 and $R^2_{10\%H2}$=0.34, all P>0.05). The expression of lincRNA-P21 decreased with time ($F_{5\%H2}$=13.54 and $R^2_{5\%H2}$=0.66, $F_{10\%H2}$=28.94 and $R^2_{10\%H2}$=0.81, all P<0.01) in both 5% and 10% H_2 concentrations groups.

2.4 Discussion

To our knowledge, this is the first study to identify the possible role of lncRNAs as biomarkers in lung cancer cells after hydrogen gas useage. Our results indicated that hydrogen not only influenced the cells' functions at the DNA and protein levels, but also affected their lncRNA expression. Furthermore, we found that LINC-PINT expression increased with increasing concentrations of H_2 gas, after 2 hours thirty minutes and three hours. Furthermore, similar to LINC-PINT, lincRNA-P21 expression was found to be lowest after using 5% H_2 gas, followed by the 10% , then 0% H_2, with

period of 2 hours and 30 minutes.

The TP53-induced transcript, LINC-PINT has been detected in multiple types of human tissue (1). LINC-PINT negatively modulates TP53 in an autoregulative manner by acting as a regulator of cell cycle arrest and a pro-survival molecule when DNA damage occurs. LincRNA-P21 is also a tumor suppressor (2), which has been shown to competitively bind to Mouse double minute 2 (MDM2), to increase the transcriptional activity of TP53 (3). Thus, LINC-PINT and lincRNA-P21 directly and indirectly, regulate cell proliferation, migration, apoptosis, and the Warburg effect (4) and they are also essential for cell growth and proliferation (5).

Hence, our findings indicate that H2 gas upregulated the expression of LINC-PINT and lincRNA-P21 in lung cancer cells, which was also reported previously (6). For example, a patient-based study involving LC patients indicated that, compared to normal tissue, the expression of lincRNA-P21 was decreased in tumor tissue (7). This downregulation of lincRNA-P21 in NSCL as also be described in Samaneh Talebi's study (8). Dongchang Wang et al reported that treatment with H2 gas inhibited the growth, migration, invasion, and apoptosis of A549 and H1975 cells by down-regulating a regulator for chromosome condensation (9).

Interestingly, however, we found that in the cells treated with H_2 gas for the longest period of time (3 hours and 40 minutes), the expression levels of LINC-PINT decreased. Our investigation also demonstrated that the expression of lincRNA-P21 decreased with time in the 5% and 10% H_2 treatment groups. Our findings are supported by Castellano et al. Their sutdy reported that lung cancer patients with a worse prognosis had higher lincRNA-P21 levels than those with a better prognosis (7).

LincRNA-p21 has been shown to be a hypoxia-responsive lncRNA that plays an important role in glycolysis by binding to HIF-1α and VHL under hypoxic circumstances (10). The Warburg effect can be defined as a form of disrupted glucose metabolism, with an increased rate of glucose consumption and production of lactate

despite the presence of oxygen, which is typical for tumors and malignant evolution (11). This metabolic characteristic the Warburg effect has also been shown to contribute to the invasion and metastasis of lung cancer malignancies (12,13). In previous studies, LINC-PINT has been found to be negatively correlated to HIF-1α, an oxygen-sensing transcription factor, in gastric cancer cells (14). Thus, it is highly probable that LINC-PINT and lincRNA-P21 are also involved in the disruption of glucose metabolism in lung cancer cells (15). Based on these data, therefore, we hypothesize that treatment with hydrogen may have induced the Warburg effect in lung cancer cells, which in turn resulted in the decreased expression of LINC-PINT and lincRNA-P21. Further research is warranted, however, to verify our hypothesis.

2.5 Conclusions

In summary -although LINC-PINT and lincRNA-P21 levels decreased in the relatively long H_2 groups- our study indicated that the expressions of LINC-PINT and lincRNA-P21 were upregulated with increasing concentrations of H_2 gas after both two hours and thirty minutes and three hours of treatment.

Thus, it can be concluded that hydrogen gas upregulated the expression of LINC-PINT and lincRNA-P21 in non-small cell lung cancer cells after a comparatively short usage period.

2.6 Limitation

Although each experiment was repeated three times, the number of repetitions may still constitute a main limitation of this study. Secondly, the H_2 treatment time ranged between 2 hours 30 minutes and 3 hours 40 minutes, which cannot be considered long time intervals. Thus, our results may have differed if longer treatment times had been applied.

3 The example of cancer risk assessment - The relationship between single nucleotide polymorphisms and skin cancer susceptibility

3.1 Introduction

Risk assessment is the earliest evidence of impending cancer in persons who don't have cancer. Single nucleotide polymorphisms (SNPs) interfere with the function of certain genes and thus may influence the probability of skin cancer (SC). The correlation between SNPs and skin cancer lacks statistical power, however. Therefore, the purpose of this study was to identify the gene polymorphisms involved in skin cancer susceptibility using network meta-analysis, and to determine the relationship between SNPs and SC risk.

3.2 Method

PubMed, Embase and Web of Science were searched for articles including ‘SNP’ and different types of SC as keywords, between January 2005 and May 2022. The Newcastle-Ottawa Scale was used to assess bias judgment. Odds ratio (ORs) and their 95% confidence intervals (CI) were determined to estimate heterogeneity within and between studies. Meta-analysis and network meta-analysis were carried out to identify the SNPs associated with SC. The P-score of each SNP was compared to obtain the rank of probability. Subgroup analyses were performed by cancer type.

3.3 Results

3.3.1 Literature search results

The literature search initially identified 3,575 studies from PubMed, Embase, and Web of Science. The search was ended on 2nd May, 2022. We screened 368 studies based on titles and abstracts and 232 full-text manuscripts. 59 studies met the inclusion criteria and were included in our network meta-analysis. One article was excluded due to bias,

as explained below.

3.3.2 Characteristics and Bias of Enrolled Studies

60 studies were published between 2005 and 2022. Studies investigating Caucasian or Mongoloid ethnicities were included. Any studies with NOS scores lower than five stars were excluded. Finally, there were 59 articles included in the systematic review and meta-analysis.

3.3.3 Pairwise meta-analysis

A direct meta-analysis was performed to determine the correlation between 275 SNPs and SC risk. 72 SNPs from 47 studies were closely associated with SC in the studies using the alleles model (A vs. B), while a significant association was found for 52 SNPs from 31 studies using the dominant model (AA+AB vs. BB). Furthermore, based on the recessive model (AA vs. AB+BB), 77 SNPs from 35 studies were related to SC. The detected SNPs were analyzed further for diagnostic accuracy.

According to the SUCRA, the allele model can be employed for exploring dominance. Then, we chose the dominant model as the genotyping model for diagnosing SC.

3.3.4 The allele model (A vs. B)

In the allele model, the major alleles of rs16891982 (G vs. C, combined OR [cOR]=2.74, 95% CI [2.20, 3.40]), rs885479 (G vs. A, cOR=1.46, 95% CI [1.06, 2.01]), rs1544410 (G vs. A, cOR=1.19, 95% CI [1.06, 1.34]), rs731236 (T vs. C, cOR=1.11, 95% CI [1.00, 1.23]), and the minor alleles of rs25487 (G vs. A, cOR=0.92, 95% CI [0.85, 0.99]), rs4911414 (G vs. T, cOR=0.85, 95% CI [0.75, 0.96]), rs1695 (W vs. M, cOR=0.79, 95% CI [0.65, 0.95]), and rs2228570 (wild-type allele vs. mutant allele, cOR=0.79, 95% CI [0.71, 0.88]) were related significantly to SC in at least two of the studies. The pooled P value for all SNPs was less than 0.05.

3.3.5 The dominant model (AA+AB vs. BB)

The results show that those who were homozygous and heterozygous for the major alleles: rs16891982 (GG+GC vs. CC, cOR=3.72, 95% CI [1.66, 8.35]), rs494379 (TT+TC vs. CC, cOR=2.62, 95% CI [1.96, 3.49]), rs514921 (AA+AG vs. GG, cOR=2.14, 95% CI [1.67, 2.75]), rs1144393 (AA+AG vs. GG, cOR=1.48, 95% CI [1.19, 1.84]), rs11615 (AA+AG vs. GG, cOR=1.41, 95% CI [1.02, 1.95]), and rs498186 (TT+TG vs. GG, cOR=1.35, 95% CI [1.10, 1.65]) had a higher risk for developing SC, than those homozygous for the minor alleles. In contrast, individuals homozygous for the minor alleles rs25487 (GG+GA vs. AA, cOR=0.85, 95% CI [0.72, 1.00]) and rs1805007 (CC+CT vs. TT, cOR=0.42, 95% CI [0.19, 0.91]) were significantly associated with increased susceptibility to SC.

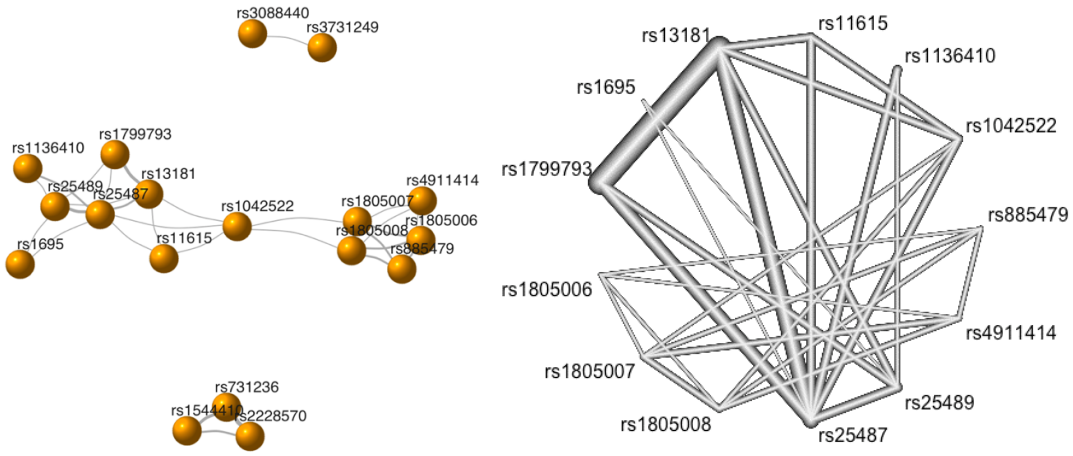
3.3.6 Subgroup analysis

Covariate regression analysis was performed for each of the three genotypes. The results showed that cancer type was not the source of heterogeneity in the studied models.

3.3.7 Network evidence

3.3.7.1 The allele model

The network plot depicts the rough comparison of each pair of SNPs (Figure 1). A node indicates an SNP, and its size represents the number of studies. The connections between the nodes mean a pair of comparisons and their thickness represents the number of direct comparisons. As is evident from Figure 1a, there were three subgroups without any connections. Also, to avoid redundancy, the network of SNPs from one study was deleted in our study. Thus, the NMA of the allele model was divided into two groups: subgroup one (including rs1544410, rs2228570, and rs731236) and subgroup two (including rs1042522, rs1136410, rs11615, rs13181, rs1695, rs1799793, rs1805006, rs1805007, rs1805008, rs25487, rs25489, rs4911414, and rs885479) (Figure 1b).



a. Network plot of SNPs in all subgroup b. Network plot of SNPs in subgroup 2

Figure 1 The network evidence plot of single nucleotide polymorphisms (SNPs) in the allele model (A vs. B).

The SNPs rs731236 vs. rs2228570 had the strongest negative correlation with SC risk in subgroup one (standardized mean differences (SMD) of OR=-0.08, 95% CI [-0.18, 0.02]). However, the P values of the correlations between the SNPs in subgroup one were above 0.05.

Similarly, the comparison with the highest direct pooled effect size in subgroup two was rs4911414 vs rs1805006 (SMD of OR=-2.94, 95%CI [-2.48, -3.40]), followed by comparison rs13181 vs. rs25489 (SMD of OR=-2.35, 95% CI [-2.54, -2.16]).

Additionally, in subgroup two, the direct and indirect evidence showed negative correlations in the comparisons of rs1042522 vs. rs25487, rs1136410 vs. rs25489, rs11615 vs. rs13181, rs11615 vs. rs25487, rs13181 vs. rs1799793, rs13181 vs. rs25487, rs1805007 vs. rs1805006, and rs1805007 vs. rs885479. However, since the indirect evidence proportion of each comparison (i.e., the mean path length of each estimated comparison) was less than 2 (16) , each of the above mentioned comparisons followed the direction of direct evidence.

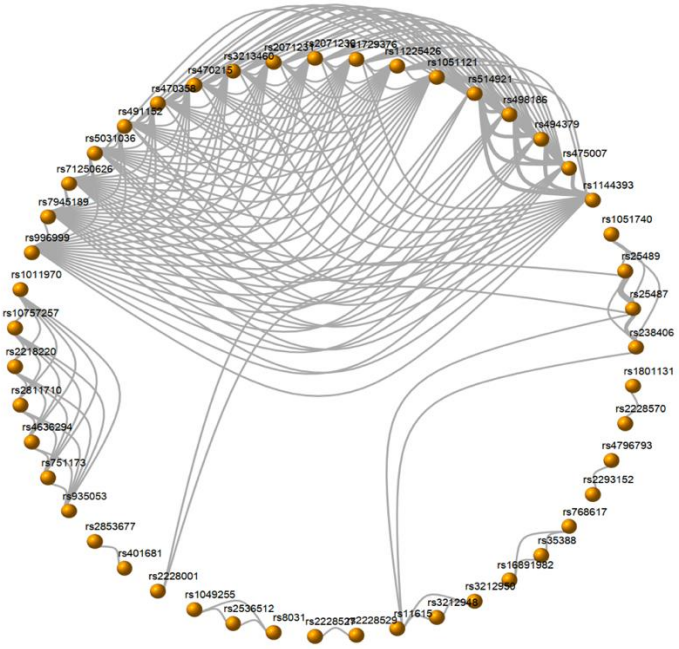
To select the SNPs with the highest chance of a significant association with skin cancer, the P scores were ranked, as shown in Table 1. The SNP rs2228570 (P-score=0.85) ranked first in subgroup one in the allele model and SNP rs13181 had the highest P-score in subgroup two (P-score=0.94).

Table 1 The rank of P-score of the SNPs in each subgroup in the Allele model

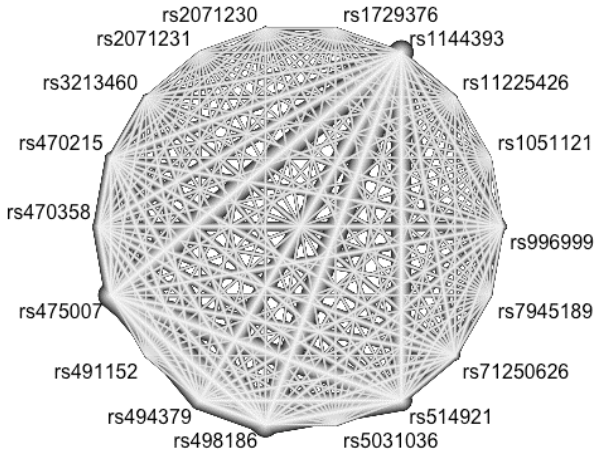
Rank	Subgroup 1	P-score	Subgroup 2	P-score
1	rs2228570	0.85	rs13181	0.94
2	rs1544410	0.47	rs1799793	0.90
3	rs731236	0.18	rs25487	0.88
4			rs11615	0.77
5			rs1042522	0.64
6			rs1695	0.57
7			rs4911414	0.54
8			rs1136410	0.41
9			rs1805007	0.33
10			rs1805008	0.24
11			rs25489	0.19
12			rs885479	0.08
13			rs1805006	0.00

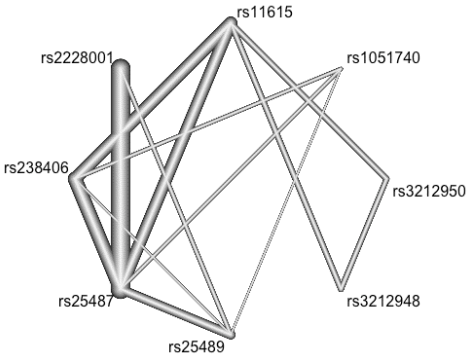
3.3.7.2 The dominant model

In Figure 2a, only two subgroups met the requirements for the NMA. Subgroup one included rs1051121, rs11225426, rs1144393, rs1729376, rs2071230, rs2071231, rs3213460, rs470215, rs470358, rs475007, rs491152, rs494379, rs498186, rs5031036, rs514921, rs71250626, rs7945189, and rs996999 (Figure 2b), while subgroup two included rs1051740, rs11615, rs2228001, rs238406, rs25487, rs25489, rs3212948, and rs3212950 (Figure 2c).



a. Network plot of SNPs in all subgroups





b. Network plot of SNPs in subgroup 1 c. Network plot of SNPs in subgroup 2

Figure 2 The network evidence plot of single nucleotide polymorphisms (SNPs) in the dominant model (AA+AB vs. BB). a. network map with 47 SNPs; b. Supgroup one network map with 18 SNPs; c. Supgroup two network map with 8 SNPs.

There was no inconsistency between the direct and indirect evidence in group one. The strongest positive correlations in this subgroup, were the comparison of rs475007 vs. rs1729376 and the comparison of rs475007 vs. rs2071231 (both SMD of network

OR=4.23, 95% CI [2.19, 6.25]). These were followed by the rs475007 vs rs491152 comparison, which SNPs were negatively correlated with SC risk (SMD of network OR=-4.21, 95% CI [−6.24, −2.18]). The comparison of rs494379 and rs514921 showed the strongest indirect correlation (SMD of indirect OR=11.52, 95% CI [-9.40, 32.44]).

The direction of direct evidence and indirect evidence were different in the comparisions of rs1144393 vs rs1051121, rs11225426 vs rs1144393, rs1144393 vs rs1729376, rs1144393 vs rs2071230, rs1144393 vs rs2071231, rs1144393 vs rs3213460, rs1144393 vs rs470215, rs1144393 vs rs470358, rs1144393 vs rs491152, rs1144393 vs rs5031036, rs1144393 vs rs71250626, rs1144393 vs rs7945189, rs1144393 vs rs996999, rs470215 vs rs514921, rs470358 vs rs498186, rs475007 vs rs514921, rs498186 vs rs514921, rs514921 vs rs71250626. However, because their indirect evidence proportion of each comparison was less, followed the direction of direct evidence.

In the subgroup two, direct and indirect evidence inconsistencies were found in the comparison of rs2228001 vs. rs25487 and the comparison of rs2228001 vs. rs25489. While the percentage of direct evidence of both these two comparisons were large than the indirect evidence. Hence, both rs25487 and rs25489 negatively correlated with rs2228001 after the network analysis. The rs238406 vs. rs25489 comparison had the strongest relationship (SMD of network OR=-2.17, 95% CI [−2.72, −1.61]).

As shown in Table 2, rs475007 has the highest P-score (0.97) in subgroup one and rs238406 has the highest P-score (0.97) in subgroup two. Therefore, the top five SNPs most likely associated with skin cancer in descending order, in subgroup one, are: rs475007, rs470358, rs498186, rs1144393, rs470215, and in subgroup twoare : rs238406, rs2228001, rs25487, rs11615, rs3212950.

Table 2 The rank of P-score of the SNPs in each subgroup

Rank	Subgroup 1	P-score	Subgroup 2	P-score
1	rs475007	0.97	rs238406	0.97

2	rs470358	0.92	rs2228001	0.87
3	rs498186	0.89	rs25487	0.62
4	rs1144393	0.84	rs11615	0.50
5	rs470215	0.79	rs3212950	0.41
6	rs514921	0.68	rs3212948	0.41
7	rs71250626	0.62	rs1051740	0.21
8	rs494379	0.59	rs25489	0.02
9	rs996999	0.58		
10	rs3213460	0.42		
11	rs2071230	0.27		
12	rs7945189	0.27		
13	rs11225426	0.26		
14	rs5031036	0.26		
15	rs1051121	0.17		
16	rs491152	0.16		
17	rs1729376	0.16		
18	rs2071231	0.16		

3.4 Discussion

Based on direct comparisons of pairwise meta-analysis and added indirect comparisons, our study employed the network meta-analysis to compare the associations between single-nucleotide polymorphisms and skin cancer using the allele model and the dominant model. Our network meta-analysis identified two subgroups in each genetic model, respectively. We ranked SNPs based on their P-score to select the most appropriate SNPs. Our results showed, that the minor alleles (T) of rs2228570 (FokI) and (C) of rs13181(ERCC2) were the highest-ranking SNPs, in both subgroups one and two, in the allele model. On the other hand, using the dominant model, the wildtype and heterozygous alleles (AA+AT) of rs475007 in subgroup one and the mutated homozygous allele (AA) of rs238406 in subgroup two were most likely to be associated

with skin cancer.

The single-nucleotide polymorphism rs2228570 (FokI) is located in the vitamin D receptor (VDR) gene. It is one of the common human VDR SNPs along with rs1544410(BsmI), rs7975232 (ApaI) and rs731236 (TaqI). Vitamin D is metabolized to vitamin D: 1,25(OH)2D3.1 in response to ultraviolet B (UVB) radiation. This metabolite is the ligand of the VDR, which in turn initiates a series of biological responses in bone metabolism, immunity, cell proliferation, and differentiation by binding to vitamin D response elements in the DNA (17). Hence, rs2228570 has not only been associated with various skin diseases, such as chronic spontaneous urticaria (CSU) (18), atopic dermatitis (AD) (19), and leprosy (20), but has also been linked to an increased incidence risk and worse prognosis of different cancers, such as breast cancer (21), ovarian cancer (22), gastric cancer (23), hepatocellular carcinoma (24), papillary thyroid cancer (25), pancreatic cancer (26) and melanoma. Our results are consistent with previous studies using assay methods (27). For instance, the study results of Zeljic et al, who used the assay method showed that the mutated genotype of rs2228570 was related to increased melanoma risk compared to the wildtype genotype in the Caucasian population (27). However, no association was observed between rs2228570 and melanoma in this investigation using the biosystem assay method (28).

SNPs rs13181 and rs238406 ranked first and second in subgroups two in both the allele and the dominant models. Both SNP alleles are located in the ERCC2 (formerly called XPD) gene. The ERCC2 polymorphisms have an ATP-dependent DNA helicase activity, which may impact DNA repair functions. Deficiency of ERCC2 has been reported to lead to xeroderma pigmentosum (XP), trichothiodystrophy (TTD), and Cockayne's syndrome (CS) (29). This observation may explain why rs13181 and rs238406 were found to be linked to cancers, such as lung cancer (30), cervical cancer (31), breast cancer, squamous cell carcinomas of the head and neck (32), and bladder cancer (33). In line with these findings, our results showed that the minor allele (C) of rs13181 and the mutated homozygous allele (AA) of rs238406 were significantly associated with SC risk. The study by Kertatbs et al. reported high frequency of the wild type allele of

rs13181 in advanced melanoma (34). However, an investigation using the microarray chip method including 1,391 NMSC cases and 2,586 cancer-free controls did not find significantly increased risks of NMSC for wildtype rs13181 (35). Furthermore, a meta-analysis found that the mutated homozygous allele (AA) of rs238406 was positively associated with the increased risk of cancer of the nervous system, the digestivetract , the genito-urinary system, and the respiratory system, but without basal cell cancer (36).

The matrix metalloproteinases (MMPs) are a family of proteolytic enzymes that are involved in cell mobility, proliferation, differentiation, and apoptosis by degrading extracellular proteins (37). MMP1, a secreted enzyme that cleaves fibrillar collagen, has been linked to cancer, by promoting cancer cell proliferation, tumor angiogenesis and vasculogenesis (38). In the dominant model of our research, all the SNPs from subgroup one were located in the MMP1 gene, and the SNP most likely to be associated with SC was rs475007. Furthermore, homozygosity for the minor allele of rs475007 was found to decrease the risk of skin cancer. Similar results were found in Hongliang Liu's study, which reported that patients homozygous or heterozygous for the major allele of rs475007 were more likely to have larger skin tumors (39).

3.5 Limitation

Due to technical differences and differences in sensitivity, our analysis only included studies that used the PCR genotypic detection method and excluded the microarray detection or genome-wide association studies (GWAS). However, GWAS allow for much larger sample sizes than PCR studies. Additionally, due to the limitations of the RStudio and StataSE softwares and the complexity of multi-arm studies, SNPs only reported in one single article were not included in the final network meta-analysis.

3.6 Future prospective

Our article indicated that people with mutations in the genes FokI (rs2228570), ERCC2 (rs13181), MMP1(rs475007) and ERCC2 (rs238406) were more likely to have skin

cancer. Dysplastic nevi (also called atypical moles) are precursors and risk factors for malignant melanoma (40). However, it is difficult to distinguish them from melanomas because of overlapping features and lack of predictive markers(41). Thus, our results may provide a possibility for the early detection of asymptomatic skin cancer if routine genetic screening is implemented in the general population in the future. Additionally, the results of our study may also provide valuable information for decision-making when determining the best mode of therapy of SC in a patient. For instance, since FokI is a vitamin D receptor gene and vitamin D is considered to be a protective factor in certain cancers, such as skin cancer (42,43), supplementation with Vitamin D may be used as adjuvant therapy in cancer patients. Therefore, identification of SC patients with FokI gene (rs2228570) mutations is important, since these patients would not benefit from adjuvant Vitamin D therapy.

In addition, we obtained direct and indirect evidence between the SNP pairs through network analysis, which proposed the possibility of hitherto unexplored relationships between certain gene mutations. For example, ERCC2 gene mutations have been shown to indirectly increase the risk of SC (44,45), and the melanocortin receptor 1 (MC1R), which encodes melanocyte-stimulating hormone (MSH) receptors, has also been shown to be risk factor for skin cancer (46). However, surprisingly, the indirect evidence of our network meta-analysis showed that ERCC2 (rs13181) was negatively related to MC1R (1805006, 1805007, 1805008, and rs885479). Therefore, the relationship between ERCC2 and MC1R, necessitates further research to determine their role in SC development.

Finally -as added scientific value - , we applied an innovative research design by performing a network analysis of case-control studies, thus providing a fresh perspective on the NMA method. Our analysis implies, that all studies involving genetically-related diseases, whether they are cohort or case-control studies, can be used to build a network in the meta-analysis, which may then provide as valuable information for the diseases' early detection, diagnosis, staging, treatment and prognosis.

4 New results

1) The expression of LINC-PINT and lincRNA-P21 upregulation with the concentration of H_2 gas.

2) LINC-PINT expression decreased in a relatively long H_2 usage time

3) The expression of lincRNA-P21 declined with the H_2 concentration.

4) The minor alleles of rs2228570 (FokI) and rs13181(ERCC2) were associated with skin cancer.

5) Wildtype and heterozygous genotypes of rs475007 (MMP1) and the mutated homozygous genotype of rs238406 (ERCC2) were most likely to be associated with skin cancer.

5 Conclusions

The expressions of LINC-PINT and lincRNA-P21 were upregulated after H_2 gas treatment. And the SNP rs2228570 (FokI), rs13181(ERCC2), rs475007 (MMP1) and rs238406 (ERCC2) can be employed as the early biomarkers for skin cancer.

6 Acknowledgements

Words cannot express my gratitude to my best supervisor Professor Istvan Kiss for his invaluable advice and unlimited patience. The way of my research is not easy. I had to change it from building a glass catfish melanoma model to a lung cancer cell line due to various uncontrollable reasons. However, Prof. Istvan Kiss supported me with all he can, every time! He is busy with huge jobs. But he still guided me in my research, my articles, my journal submission, and my career plan every week, even using his private rest time sometimes. To my research, he applied for many times funding from different

schools and asked for help with cell lines and equipment from different universities. There are too many things he helped me with that I cannot return the favour.

Besides, I am extremely grateful to my co-supervisor Professor Hongbo Liu from China Medical University. Without her help, I also cannot finish my PhD. Thanks for her support for the methodologies, analysing and software of my research. She got nothing paid, but she shared all her knowledge and skills without reservation. Thank her for listening to, supporting, inspiring and impacting me in my life, study, career and dream over the past four years.

I am also grateful to my colleagues, István Szabó, Brunnerné Bayer Zsuzsanna, Ágnes Szenczi, Timea Varjas and Eva Pozsgai, for their kind help with my equipment, cell line, PCR, and results. They helped me again and again without any impatience. I also could not have undertaken this journey without their help.

I cannot begin to express my gratitude and appreciation for my parents, families and friends. None of these could have happened without their financial and emotional support. Their belief in me has kept my spirits and motivation high.

At last, thanks must go to the chair of my committees, for their invaluable patience and feedback. Thanks for your generous knowledge and expertise.

7 List of Publications

7.1 Thesis based articles

- The relationship between single nucleotide polymorphisms and skin cancer susceptibility: A Systematic Review and Network Meta-Analysis (IF=6.24)

Lu Zhang, Eva Pozsgai, Yongan Song, John Macharia, Huda Alfatafta, Jia Zheng, Zhaoyi Li, Hongbo Liu, István Kiss

Front Oncol (2023) 13:1094309. doi: 10.3389/fonc.2023.1094309

- The therapeutic effect of hydrogen gas on lung cancer - LINC-PINT and lincRNA-P21 as biomarkers

Lu Zhang, Timea Varjas, Eva Pozsgai, István Szabó, Ágnes Szenczi, Huda Alfatafta, Yongan Song, John Macharia, Hongbo Liu, István Kiss (IF=1.4)

Genetic Testing and Molecular Biomarkers. Submitting for publication

7.2 Other publications

- Lower handgrip strength levels probably precede triglyceride glucose index and associated with diabetes in men not in women

Jia Zheng, Lu Zhang, Min Jiang

Journal of Diabetes Investigation, vol. 13, no. 1, pp. 148–155, 2022, doi: 10.1111/jdi.13626.

- Are chemical compounds in medical mushrooms potent against colorectal cancer carcinogenesis and antimicrobial growth?

John M. Macharia, Lu Zhang, Ruth W. Mwangi, Nora Rozmann, Zsolt Kaposztas, Tímea Varjas, Miklós Sugár, Huda Alfatafta, Márton Pintér, Raposa L. Bence

Cancer Cell Int, vol. 22, no. 1, p. 379, Dec. 2022, doi: 10.1186/s12935-022-02798-2.

- Microarray data analysis to identify miRNA biomarkers and construct the lncRNA-miRNA-mRNA network in lung adenocarcinoma

Yongan Song, Leonardo Kelava, Lu Zhang, István Kiss

Medicine, vol. 101, no. 36, p. e30393, Sep. 2022, doi: 10.1097/MD.0000000000030393.

- Effect of the knee replacement surgery on activity level based on ActivPAL: a systematic review and meta-analysis study

Alfatafta Huda, Alfatafta Mahmoud, Onchonga David, Hammoud Sahar, Khatatbeh Haitham, Zhang Lu, Boncz Imre, Lohner Szimonetta, Molics Bálint

Musculoskeletal Disorders, vol. 23, no. 1, p. 576, Jun. 2022, doi: 10.1186/s12891-022-05531-2.

- Effect of using knee valgus brace on pain and activity level over different time intervals among patients with medial knee OA: systematic review

Huda Alfatafta, David Onchonga, Mahmoud Alfatafta, Lu Zhang, Imre Boncz, Szimonetta Lohner, Bálint Molics

BMC Musculoskeletal Disorders, vol. 23, no. 1, p. 576, Jun. 2022, doi: 10.1186/s12891-022-05531-2.

- Focusing on the premature death of redeployed miners in China: an analysis of cause-of-death information from non-communicable diseases

Wei Xian, Bing Han, Leizhen Xia, Yining Ma, Haodi Xu, Lu Zhang, Li Li, Hongbo Liu

Global Health, vol. 15, p. 7, Jan. 2019, doi: 10.1186/s12992-019-0450-5.

- Effect of socio-demographic characteristics on social security in northeast China

Wei Xian, Cheng Jin, Bing Han, Xueying Xu, Lu Zhang, Hongbo Liu

Health Soc Care Community, Aug. 2020, doi: 10.1111/hsc.13127.