



**HAL**  
open science

## **FGF21 signalling pathway and metabolic traits - genetic association analysis**

Maciej Tomaszewski, Bernhard M Kaess, Timothy A Barnes, Fadi J Charchar, Klaus Stark, Dawn Waterworth, Kijoung Song, Peter Vollenweider, Gerard Waeber, Vincent Mooser, et al.

► **To cite this version:**

Maciej Tomaszewski, Bernhard M Kaess, Timothy A Barnes, Fadi J Charchar, Klaus Stark, et al.. FGF21 signalling pathway and metabolic traits - genetic association analysis. *European Journal of Human Genetics*, 2010, 10.1038/ejhg.2010.130 . hal-00567033

**HAL Id: hal-00567033**

**<https://hal.science/hal-00567033>**

Submitted on 18 Feb 2011

**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.

**FGF21 signalling pathway and metabolic traits - genetic association analysis**

Bernhard M Kaess<sup>1,2</sup>, Timothy A Barnes<sup>1</sup>, Klaus Stark<sup>2</sup>, Fadi J Charchar<sup>3</sup>, Dawn Waterworth<sup>4</sup>, Kijoung Song<sup>4</sup>, William YS Wang<sup>1,5</sup>, Peter Vollenweider<sup>6</sup>, Gerard Waeber<sup>6</sup>, Vincent Mooser<sup>4</sup>, Ewa Zukowska-Szzechowska<sup>7</sup>, Nilesh J Samani<sup>1</sup>,  
Christian Hengstenberg<sup>2</sup>, Maciej Tomaszewski<sup>1</sup>

<sup>1</sup>Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; <sup>2</sup>Klinik und Poliklinik für Innere Medizin II, University of Regensburg, Regensburg, Germany;

<sup>3</sup>School of Science and Engineering, University of Ballarat, Ballarat, Australia;

<sup>4</sup>GlaxoSmithKline, King of Prussia, Pennsylvania, USA; <sup>5</sup>School of Medicine,

University of Queensland, Brisbane, Australia; <sup>6</sup>Department of Medicine, Internal

Medicine, CHUV, Lausanne, Switzerland; <sup>7</sup>Department of Internal Medicine,

Diabetology, and Nephrology, Medical University of Silesia, Zabrze, Poland

Corresponding author: Maciej Tomaszewski, Department of Cardiovascular Sciences, University of Leicester, Clinical Sciences Wing, Glenfield General Hospital, Leicester, LE3 9QP, UK, tel: +44 116 256 3028, fax: +44 116 287 5792, e-mail: mt142@le.ac.uk

Running Title: Genetic association analysis of FGF21 pathway

Key words: fibroblast growth factor 21, fibroblast growth factor receptor 2, cholesterol, single nucleotide polymorphism, association

**Word count Abstract: 249; Word count Manuscript: 4,631; Number of Tables and Figures: 3+1; Number of Supplementary Tables and Figures: 3+2**

## **Abstract**

Fibroblast growth factor 21 (FGF21) is a novel master regulator of metabolic profile. The biological actions of FGF21 are elicited upon its klotho beta (KLB)-facilitated binding to FGF receptor 1 (FGFR1), FGF receptor 2 (FGFR2) and FGF receptor 3 (FGFR3). We hypothesised that common polymorphisms in the FGF21 signalling pathway may be associated with metabolic risk. At the screening stage we examined associations between 63 common single nucleotide polymorphisms (SNPs) in 5 genes of this pathway (FGF21, KLB, FGFR1, FGFR2, FGFR3) and 4 metabolic phenotypes (LDL cholesterol - LDL-C, HDL-cholesterol, triglycerides and body mass index - BMI) in 629 individuals from Silesian Hypertension Study. Replication analyses were performed in 5,478 unrelated individuals of the Swiss CoLaus cohort (imputed genotypes) and in 3,030 directly genotyped individuals of the German Myocardial Infarction Family Study. Of 54 SNPs that met quality control criteria after genotyping in Silesian Hypertension Study, four (rs4733946 and rs7012413 in FGFR1; rs2071616 in FGFR2 and rs7670903 in KLB) showed suggestive association with LDL-C ( $p=0.0006$ ,  $p=0.0013$ ,  $p=0.0055$ ,  $p=0.011$ , respectively) and one (rs2608819 in KLB) was associated with BMI ( $p=0.011$ ); all - false discovery rate  $q<0.5$ . Of these, only one FGFR2 polymorphism (rs2071616) showed replicated association with LDL-C in both the CoLaus cohort ( $p=0.009$ ) and men from the German Myocardial Infarction Family Study ( $p=0.017$ ). The direction of allelic effect of rs2071616 upon LDL-C was consistent in all examined populations. These data show that common genetic variation in FGFR2 may be associated with LDL-C in subjects of white European ancestry.

## **Introduction**

Genetic epidemiology has made significant progress over the last few years. With the availability of genotyping platforms that can analyse hundreds of thousands of single nucleotide polymorphisms (SNPs) it has become feasible to perform high resolution scans over the whole human genome (genome-wide association studies - GWAS), thereby identifying genetic variants associated with phenotypes of interest. The major advantage of GWAS is that they do not require *a priori* assumptions about a potential role of alleles, genes and pathways in regulation of the examined trait. However, because of the scale of such analysis the problem of substantial correction for multiple testing arises and, therefore, representative cohorts and replication are needed to reduce the probability of false positive reports. Furthermore, despite the extremely large number of SNPs on the platforms currently available, many single loci are poorly covered on a majority of platforms and therefore a significant proportion of potentially relevant variants may not be directly genotyped. Thus, GWAS are a tool of choice to discover disease genes that have not been suspected before, but may have limited value in analysis of candidate genes or pathways with strong functional potential to affect the phenotype of interest. For this type of investigation, systematic exploration of all common allelic variants in genes that encode molecules within the selected pathway is the method of choice. We used this approach to examine associations of common genetic variation in the fibroblast growth factor 21 (FGF21) signalling pathway with metabolic disturbances in white European population.

The FGF21 pathway is a novel metabolic signalling cascade recently identified as a master hormonal regulator of glucose, lipids and overall energy balance <sup>1</sup>. Specifically, endogenous FGF21 was shown to facilitate insulin-independent uptake of glucose in

adipocytes <sup>2</sup> and to promote lipolysis <sup>3</sup> contributing to reduction of fat storage <sup>4</sup>. Exogenous FGF21 administration in experimental models was demonstrated to lower both glucose and triglyceride plasma levels and promote weight loss <sup>2</sup>. These effects may be mediated via 3 types of fibroblast growth factor receptor (termed FGFR1, FGFR2 and FGFR3, respectively) <sup>5-7</sup>. Activation of these FGFRs by FGF21 is conditional upon co-expression of klotho beta (KLB) <sup>5-7</sup>, a transmembrane protein that has been discovered through its structural similarity to klotho (which mediates receptor binding of FGF23).

We have hypothesised that common polymorphisms in 5 genes of FGF21 signalling pathway may be associated with determinants of metabolic risk in subjects of European origin.

## **Methods**

### Subjects

Primary association analyses were carried out in the Silesian Hypertension Study (SHS) that has been previously described <sup>8</sup>. In brief, SHS is a collection of 207 Polish Caucasian families (629 individuals, 49.8% men) recruited through essentially hypertensive offspring. Each individual was phenotyped according to a standard protocol (collection of clinical information by coded questionnaire, weight, height and blood pressure measurements, basic anthropometry, biochemical phenotyping). Blood samples were taken under fasting conditions. Total cholesterol (TC), HDL-C and triglycerides (TGs) were determined using standard enzymatic methods on a Cobas Bio-Autoanalyzer and LDL-C was calculated using the Friedewald formula. None of the subjects was treated pharmacologically with lipid-lowering medication.

Replication analyses were carried out in the Swiss CoLaus cohort and the German Myocardial Infarction Family Study (GerMIFS).

CoLaus is a random sample of 6,188 unrelated individuals (2,937 men) from the general population of Lausanne, Switzerland, and has been previously described in detail <sup>9</sup>. In short, randomly selected Caucasian Lausanne inhabitants aged 35-75 years were invited to join the study. Eligible subjects were phenotyped in the outpatient clinic of the Centre Hospitalier Universitaire Vaudois (CHUV). Phenotyping included taking clinical information, anthropometry and blood pressure measurements. Blood samples were taken after overnight fasting. LDL-C was calculated using the Friedewald formula (in subjects with triglycerides < 4.6 mmol/l) and was adjusted for lipid-lowering treatment by adding 1.5 mmol/l to measured values of subjects who were managed pharmacologically for hyperlipidaemia.<sup>10</sup>

GerMIFS was based on 1,261 families (3,030 patients; 60.6% men) of white European ancestry. The details of this study have been previously described <sup>11</sup>. In brief, families were recruited through index patients with coronary artery disease (CAD) at 15 cardiac rehabilitation centres distributed throughout Germany. All index patients had MI before 60 years of age. If at least one sibling presented with MI or severe CAD before 70 years of age, the entire nuclear family (index patient, available parents, and all affected and unaffected siblings) was contacted and invited to participate in the study. Blood was drawn from non-fasting individuals. TC, LDL-C, HDL-C and TGs were determined directly using standard enzymatic methods. For the purpose of replication study of LDL-C, we excluded all patients on lipid-lowering treatment at the time of recruitment, leaving 3,115 patients available for analysis. After excluding patients with missing

phenotypes, we were able to perform association analyses with LDL-C in 3,030 patients (1,261 families, 1,835 men).

The local Ethics Committees approved the study protocols, and all participants gave written, informed consent. All protocols conform to the principles outlined in the Declaration of Helsinki.

#### SNP selection

Tagging SNPs in FGF21, FGFR1, FGFR2, FGFR3 and KLB were selected from the HapMap database to reach a gene coverage of >90% with  $r^2 > 0.8$  for all SNPs with a minor allele frequency of  $\geq 0.1$  in the CEU population. In addition, each locus was saturated with additional common potentially functional polymorphisms (CpG islands, DNase I hypersensitive sites, microRNA target sites)<sup>12</sup>. Altogether, 63 SNPs were selected for genotyping in SHS. Of these, after genotyping, quality filters and association testing, 5 SNPs with high probability of a true association in SHS were selected for further replication.

#### Genotyping

Genotyping in SHS was carried out using iPLEX<sup>TM</sup> assays on the Sequenom MassArray system (Sequenom Inc, San Diego, CA, USA). SNPs in CoLaus cohort were retrieved from a genome-wide scan conducted using the Affymetrix Genome-Wide Human SNP Array 5.0. Specifically SNP rs2608819 was directly genotyped and the other 4 SNPs were imputed using the Quicktest imputation software (the quality metrics were: rs7012413: 0.999, rs4733946: 0.824, rs2071616: 0.302, rs7670903: 0.914). Replication genotyping of rs2071616 in GerMIFS was performed using the TaqMan assays

(Applied Biosystems, Foster City, CA, USA) on ABI PRISM 7900HT Sequence Detection System (Applied Biosystems).

#### Statistical and bioinformatic analysis

The following quality control filters were applied to SNPs before the statistical analyses: call rate  $\geq 0.9$ , no deviation from Hardy-Weinberg equilibrium (HWE) in the parental generation (HWE  $p > 0.01$ ) and number of Mendelian errors  $< 3$ .

The primary association analyses in SHS were performed using family-based association testing (*FBAT* version 2.0.2)<sup>13</sup> under the null hypothesis of no linkage and no association. Replication analyses in CoLaus were based on linear regression models constructed in SAS v.9.1 (SAS Institute Inc., Cary, North Carolina). In GerMIFS association analysis was performed using generalised estimating equation (GEE) regression with an exchangeable familial correlation structure, as implemented in the *geepack* library in R<sup>14</sup>.

We used FBAT for our exploratory primary analysis in SHS because this tends to produce less false positive results than GEE<sup>15</sup> and performs well in samples consisted of nuclear families with available parents (such as SHS). Generally, FBAT and GEE use different approaches to handle family structure: FBAT is based on Mendelian laws in families and therefore requires a certain minimum information content of the pedigree, which depends on family structure, genotypes and phenotypes. The GerMIFS cohort had initially been recruited for microsatellite linkage analyses and consists to a large extent mainly of sib-pairs. Given that about 50% of the pedigrees would not contribute to the FBAT statistic, we did not use FBAT in this cohort. GEE regression, on the other hand, is based on classical population-based association and therefore is more



susceptible to bias due to stratification and admixture effects. However, in the GerMIFS study, the vast majority of index cases has been included into GWA studies<sup>16,17</sup>, showing no signs of population substructure.

All association tests were performed assuming an additive model of inheritance. All models were adjusted for age, age<sup>2</sup> and gender (where applicable). All p-values are given for two-tailed tests.

To correct for multiple testing in the primary family-based analysis we calculated false discovery rates (FDR) q-values. FDR describe the expected number of false-positives among all positive results. Q-values were calculated using the *R*-based tool q-value (smoother algorithm,  $\lambda$  range 0.0 – 0.9)<sup>18</sup>, including all 216 performed association tests on the 4 traits analysed in SHS. At the replication stage we used Bonferroni correction.

Haplotypes and linkage disequilibrium (LD) in FGFR2 were constructed using the *Haploview* program<sup>19</sup> and visualized using the *Locusview* software. *LocusView* is a Java-based application that visualises chromosomal regions annotated with genomic features and experimental data (such as results from association analysis (<http://www.broadinstitute.org/science/programs/medical-and-population-genetics/locusview20>)). The software was developed by Petryshen, Kirby, and Ainscow and is freely available.

The graphical representation of FGF21 signalling cascade (Figure 1, panel B) was constructed using the GeneGo software ([www.genego.com](http://www.genego.com)). Sequence alignment (Figure 1, panel C) was derived from the ClustalW2 portal ([www.ebi.ac.uk/Tools/clustalw2/index.html](http://www.ebi.ac.uk/Tools/clustalw2/index.html)). Information on functional potential of SNPs was obtained *in silico* from SNP Function Annotation Portal <http://brainarray.mbni.med.umich.edu/Brainarray/Database/SearchSNP/snpfunc.aspx>.

## Results

The brief clinical and anthropometric characteristics of the three study populations are listed in Table 1. Additional characteristics of the cohorts have been described elsewhere<sup>9,11,20</sup>.

### Primary association analysis - SHS

Of 63 selected and genotyped polymorphisms 9 SNPs were excluded before the association analysis due to poor genotyping quality. The mean call rate of the remaining 54 SNPs (Supplementary Table 2) was 97%. At  $r^2=0.8$ , we reached a good tagging coverage of all HapMap SNPs in our candidate genes (100% for FGF21, 96% for FGFR1, 90% for FGFR2, 100% for FGFR3 and 78% for KLB). Under additive model of inheritance primary family-based analysis revealed 14 nominal genetic associations with metabolic phenotypes in SHS – 6 with BMI (rs1982738, rs7674434, rs12152703, rs17431867, rs2608819, rs3135859), 6 with LDL-C (rs4733946, rs7012413, rs3135766, rs3135763, rs2071616, rs7670903), 1 with HDL-C (rs2608819) and 1 with TGs (rs3135761) (Table 2). Of those, only 5 (rs7012413, rs4733946, rs2071616, rs7670903 with LDL-C, rs2608819 with BMI) showed q-values  $<0.5$  (suggestive associations) (Table 2). The details of this analysis are given in Supplementary Table 2.

### Replication

5 SNPs (rs7012413, rs4733946, rs2071616, rs7670903, rs2608819) with q-values  $< 0.5$  were taken forward for *in silico* replication in the CoLaus cohort. Of those, FGFR2 rs2071616 polymorphism showed a significant association with LDL-C ( $p=0.009$ , Table 3), even after correcting for multiple testing (Bonferroni correction threshold: 0.01;

0.05/5). The direction of the allelic effect on LDL-C was consistent with the findings from SHS – in both cohorts the minor allele was associated with higher LDL-C. The quantitative analysis in CoLaus showed that each minor allele copy of rs2071616 increased LDL-C by approximately 0.092 mmol/l ( $\beta= 0.092$ , SE=0.004).

We genotyped rs2071616 in the third independent cohort - the GerMIFS study and found that rs2071616 was associated with LDL-C amongst men only ( $p=0.017$ ), but not in the entire cohort of subjects ( $p=0.32$ ). However, both the direction of allelic association and the magnitude of minor allele effect of rs2071616 on LDL-C in GerMIFs men were strikingly similar to CoLaus subjects ( $\beta=0.094$ , SE=0.039).

## **Discussion**

The most important finding from our analysis is a suggestive association between a genetic variant of FGFR2 (SNP rs2071616) and LDL-C in white European subjects. This association was identified through a systematic screening of all common variants in 5 genes of the FGF21 signalling cascade in Polish families and was replicated in subjects from the general Swiss population and German men from families with clustering of CAD. To the best of our knowledge, no associations between FGFR2 and lipids have been reported by GWAS or candidate gene studies.

The selection of FGF21 signalling cascade as a novel candidate pathway of metabolic regulation is supported by multiple lines of evidence from experimental models. The central molecule of this pathway - FGF21 - promotes glucose uptake into adipocytes and improves glucose tolerance in diabetic rodents without inducing hypoglycemia <sup>2</sup>. Furthermore, mice overexpressing FGF21 have lower triglyceride plasma levels and are less prone to diet-induced obesity <sup>2</sup>. When administered to diabetic monkeys, FGF21

improves glucose tolerance (without inducing hypoglycemia), lowers LDL-C and elevates HDL-C<sup>21</sup>. Knockdown of FGF21 in mice leads to rapid hepatic fat accumulation and prominent lipid disturbances<sup>22</sup>. Acting downstream in the PPAR $\alpha$  signalling cascade, FGF21 regulates metabolic adaptation to the fasting state. Under a ketogenic diet or in the fasting state, FGF21 stimulates lipolysis in white adipose tissue and hepatic ketogenesis<sup>3,4,22</sup>. These metabolic effects of FGF21 become apparent upon activation of at least 3 types of FGF receptor (FGFR1, FGFR2 and FGFR3) along with transmembrane mediator protein KLB<sup>5-7</sup>. The further signalling cascade has been shown to involve PPAR $\gamma$ -coactivator protein 1 $\alpha$ <sup>23</sup>. Interestingly, both the FGF21-induced tyrosine phosphorylation activity of FGFR2 and circulating FGF21 levels are augmented by the PPAR $\gamma$ -ligand rosiglitazone<sup>24,25</sup>, suggesting a feedback mechanism in metabolic regulation.

Our most significant association signal maps to FGFR2 (Figure 1) – a molecule known as a critical regulator of cell proliferation. Multiple splice variants of FGFR2 are expressed in many tissues, including vasculature and liver<sup>26</sup>, and bind to different FGF ligands including FGF21. The role of FGFR2 in metabolism is primarily determined by tissue-specific co-expression of KLB<sup>4</sup>: FGF receptors bind to the N-terminus of FGF21, while KLB interacts with the C-terminus<sup>27</sup>. Genetic variation within FGFR2 has been linked to a wide range of phenotypes - from congenital syndromes of complex craniofacial malformations such as Crouzon syndrome (Dysostosis craniofacialis), Apert syndrome (acrocephalosyndactyly) and craniosynosthosis to breast cancer and other malignancies such as gastric, lung, ovarian and endometrial cancer<sup>28</sup>. The FGFR2 syntenic region in rats is located within a quantitative trait locus (QTL) for hypertriglyceridemia, type 2 diabetes and body weight ([www.ucsc.org](http://www.ucsc.org)), but, to the best

of our knowledge, associations between FGFR2 mutations and LDL-C have not been reported before.

The rs2071616 polymorphism is situated in intron 6 (approximately 100 bp from the nearest exon-intron boundary) of FGFR2 (isoform 2 precursor). The polymorphism is located within a DNA segment that shows a high level of conservation in mammalian evolution (Figure 1, panels C-D). There are not any known functional polymorphisms that could account for this association through high ( $r^2 \geq 0.8$ ) LD with rs2071616 (Supplementary Table 3, Supplementary Figure 1). The only statistically similar SNP (rs2912762 - in complete LD with rs2071616 in HapMap CEU individuals) maps to another intronic segment of FGFR2 and does not appear to exhibit any obvious biological potential (Supplementary Table 3) but is conserved (although to a lesser extent than rs2071616) in mammals. Further studies will be needed to clarify whether rs2071616, its proxy (rs2912762) or in fact an unidentified functional polymorphism in LD with rs2071616, is an actual driver of the association.

We should acknowledge several limitations of the study - first, unavailability of plasma glucose levels in SHS did not permit us to include this metabolic phenotype in the primary association analysis. Second, the replication studies in CoLaus cohort were based on data generated using imputation algorithms and rs2071616 was at the bottom end of what has been accepted as a valid imputation <sup>29</sup>. Third, our cohorts were recruited by different strategies: SHS for hypertension, GerMIFS for familial clustering of CAD, CoLaus from general population. Fourth, we replicated our main association finding (rs2071616 and LDL-C) in German men only and not in women. Whether gender-specific effects may account (at least to some extent) for these findings will

require further investigations in other cohorts. Finally, we should also acknowledge the limitations of multiple testing. We analysed over 50 SNPs and 4 different phenotypes, resulting in over 200 tests in SHS. A conventional adjustment for multiple testing such as Bonferroni correction would have been overly conservative given that neither genotyped SNP nor examined phenotypes were completely independent. Indeed, many SNPs were in LD with each other and lipid fractions showed high levels of correlation. Although calculation of false discovery rate (q-values) does not account for this LD-driven or phenotypic non-independence we used a non-conservative threshold of 50% ( $q\text{-value} < 0.5$ ) to identify suggestive associations. The quantile-quantile plot of  $-\log p$ -values from all 216 tests performed in SHS revealed deviation in the extreme tail of association statistics distribution from the null distribution indicating that the findings above the point of the deviation (mainly our suggestive associations) could indeed represent genuine results (Supplementary Figure 2).

## **Conclusions**

In summary, the identified suggestive genetic association between FGFR2 and LDL-C illuminates a novel potential mechanism of cholesterol regulation. We also demonstrate that systematic pathway-oriented genetic analyses may contribute to gene discovery in the era of GWAS. Further replication and functional analyses of FGFR2 in a metabolic context are warranted to validate the association identified through genetic screening of the relevant signalling pathway.

**Acknowledgements**

This project was supported by NIH Fogarty International Research Collaboration Award (R03 TW007165; to M.T.). BK's fellowship in Leicester was supported by the EU project Cardiogenics (LSHM-CT-2006-037593). T.B. was supported by a British Heart Foundation project grant (PG/06/097; to M.T.). N.J.S. holds a British Heart Foundation Chair of Cardiology. The CoLaus study was supported by research grants from GlaxoSmithKline and from the Faculty of Biology and Medicine of Lausanne, Switzerland, and is currently supported by the Swiss National Science Foundation (grant no: 33CSCO-122661). We are grateful to Dr Allan Beveridge for support in bioinformatic analysis.

**Disclosure**

No conflicts of interest to disclose.

Table 1: General clinical characteristics of the study cohorts

Phenotype	<b>SHS</b>	<b>CoLaus</b>	<b>GerMIFS</b>
No. of subjects	629	5,478	3,030
No. of families	207	0	1,261
No. of men (%)	313 (49.8)	2,612 (47.7)	1,835 (60.6)
Age, years	45.8 ± 15.7	53.2 ± 10.8	59.1 ± 10.7
BMI, kg/m <sup>2</sup>	26.8 ± 4.7	25.8 ± 4.6	27.2 ± 3.9
LDL-cholesterol, mmol/L	4.0 ± 1.10	3.47 ± 1.01	4.17 ± 1.14
HDL-cholesterol, mmol/L	0.98 ± 0.29	1.63 ± 0.44	1.36 ± 0.38
Triglycerides, mmol/L	1.76 ± 1.10	1.40 ± 1.21	2.19 ± 1.42 *
Lipid-lowering medication (%)	0	472 (8.6)	excluded

Data are means ± standard deviations or counts and percentages.

SHS: Silesian Hypertension Study; CoLaus: population-based cohort of Lausanne;

GerMIFS: German Myocardial Infarction Family Study. \*: non-fasting.



Table 2: Nominally significant associations between SNPs of FGF21 signalling pathway and metabolic phenotypes in primary age-, age<sup>2</sup>- and gender-adjusted family-based analysis (SHS).

<b>Trait</b>	<b>Gene</b>	<b>SNP</b>	<b>Alleles</b>	<b>MA</b>	<b>MAF</b>	<b>Risk allele</b>	<b>Z-value</b>	<b>p-value</b>	<b>q-value</b>
<b>BMI</b>	FGFR3	rs3135859	G/T	T	0.22	G	2.13	0.033	0.68
	KLB	rs1982738	A/G	T	0.21	C	2.04	0.041	0.68
		rs7674434	G/T	G	0.29	G	2.30	0.021	0.59
		rs12152703	G/T	T	0.23	T	2.11	0.035	0.68
		rs17431867	C/T	T	0.18	T	1.98	0.048	0.68
		rs2608819	C/T	T	0.11	C	2.55	0.011	0.42
<b>LDL-C</b>	FGFR1	rs7012413	C/T	T	0.30	T	3.21	0.0013	0.14
		rs4733946	G/T	T	0.09	T	3.42	0.0006	0.13
	FGFR2	rs3135766	C/T	G	0.45	G	2.09	0.037	0.68
		rs3135763	G/T	A	0.18	A	2.05	0.041	0.68
		rs2071616	C/T	T	0.35	T	2.78	0.0055	0.39
	KLB	rs7670903	A/G	A	0.38	A	2.57	0.011	0.47
<b>HDL-C</b>	KLB	rs2608819	C/T	T	0.11	C	-2.24	0.025	0.68
<b>triglycerides</b>	FGFR2	rs3135761	C/T	A	0.23	G	2.39	0.026	0.68

SHS - Silesian Hypertension Study, BMI – body mass index, LDL-C – LDL-cholesterol, HDL-C – HDL-cholesterol, SNP – single nucleotide polymorphism, MA –

minor allele, MAF – minor allele frequency in the parental generation of Silesian Hypertension Study, Z-value – family-based association test that illustrates the magnitude of the genetic association and the direction of the allelic association (“-“ refers to the decrease in the quantity of the trait), p-value – level of statistical significance of family-based association test, q-value – false discovery rate.

Table 3: Replication of suggestive associations identified in Silesian Hypertension Study in CoLaus cohort.

<b>Trait</b>	<b>Gene</b>	<b>SNP</b>	<b>MA</b>	<b>MAF</b>	<b>Risk allele</b>	<b><math>\beta</math>-coefficient</b>	<b>SE</b>	<b>p-value</b>
LDL-C	FGFR1	rs7012413	T	0.30	T	0.018	0.021	0.37
		rs4733946	T	0.08	T	0.027	0.037	0.48
	FGFR2	rs2071616	T	0.31	T	0.092	0.035	0.009
	KLB	rs7670309	A	0.42	G	0.024	0.019	0.21
BMI	KLB	rs2608819	T	0.13	T	0.014	0.128	0.91

BMI – body mass index, LDL-C – LDL-cholesterol, SNP – single nucleotide polymorphism, MA – minor allele, MAF– minor allele frequency,  $\beta$ -coefficient - estimated quantitative effect of each SNP minor allele copy on a metabolic phenotype (adjusted for age, age<sup>2</sup> and sex), SE – standard error

Figure 1: Fibroblast growth factor receptor 2 gene (FGFR2) - association and evolutionary analysis

Panel A, Upper:  $-\log$  transformed probability values from the analysis of association between FGFR2 and LDL-C in Silesian Hypertension Study; Lower: Linkage disequilibrium (LD) map ( $r^2$  coefficient-based) of FGFR2 SNPs. Note: Dark red squares indicate high  $r^2$  and therefore high LD, white squares indicate  $r^2=0$  and no LD.

Panel B: FGFR2 binds FGF21 in the presence of KLB (a transmembrane co-factor); B-binding.

Panel C: Sequence alignment of region containing SNP rs2071616 in FGFR2 across 15 mammalian species, position of rs2071616 is marked in red.

Panel D: Evolutionary conservation of region in FGFR2 gene surrounding SNP rs2071616, position of which is marked by arrow.

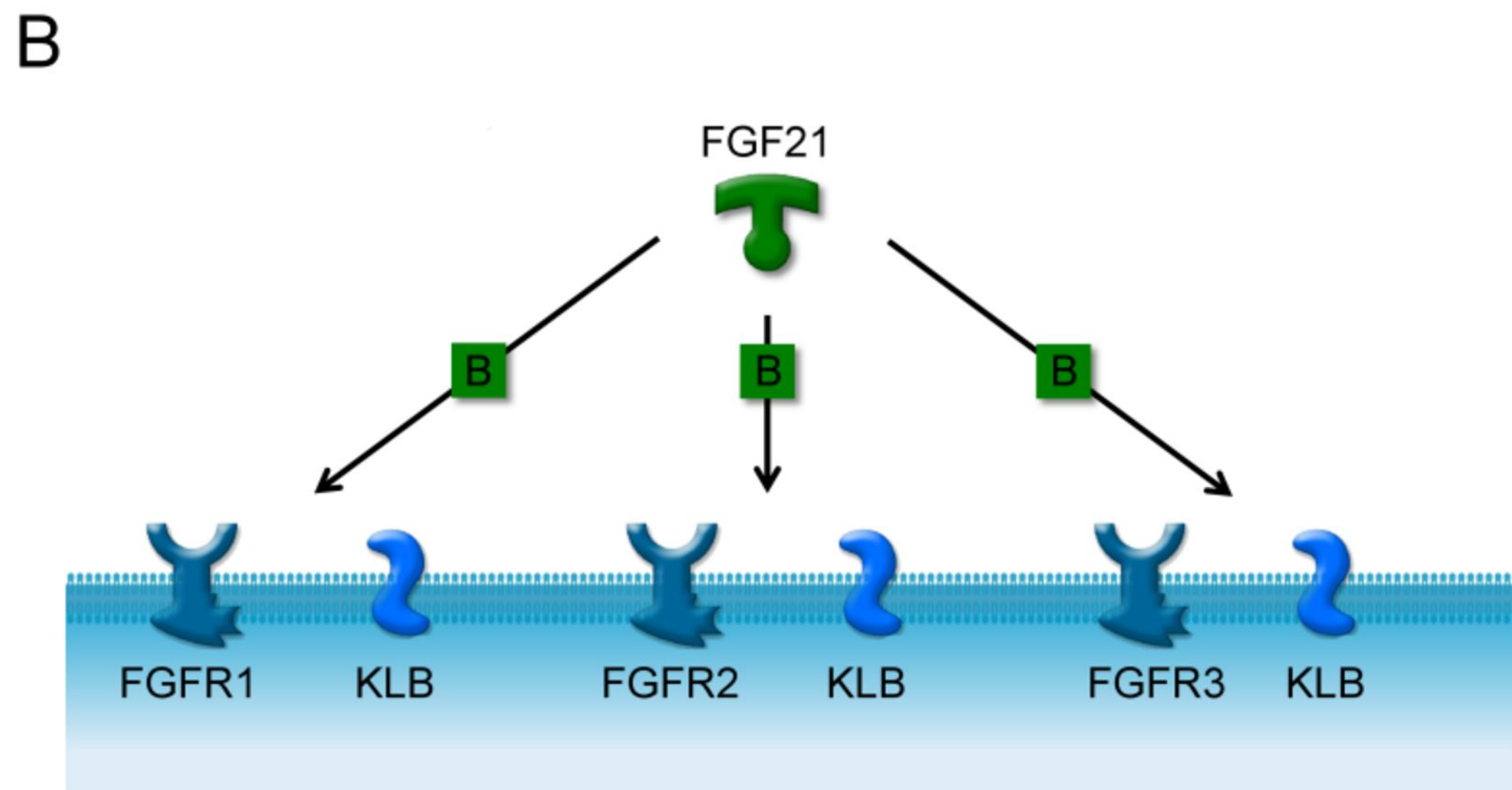
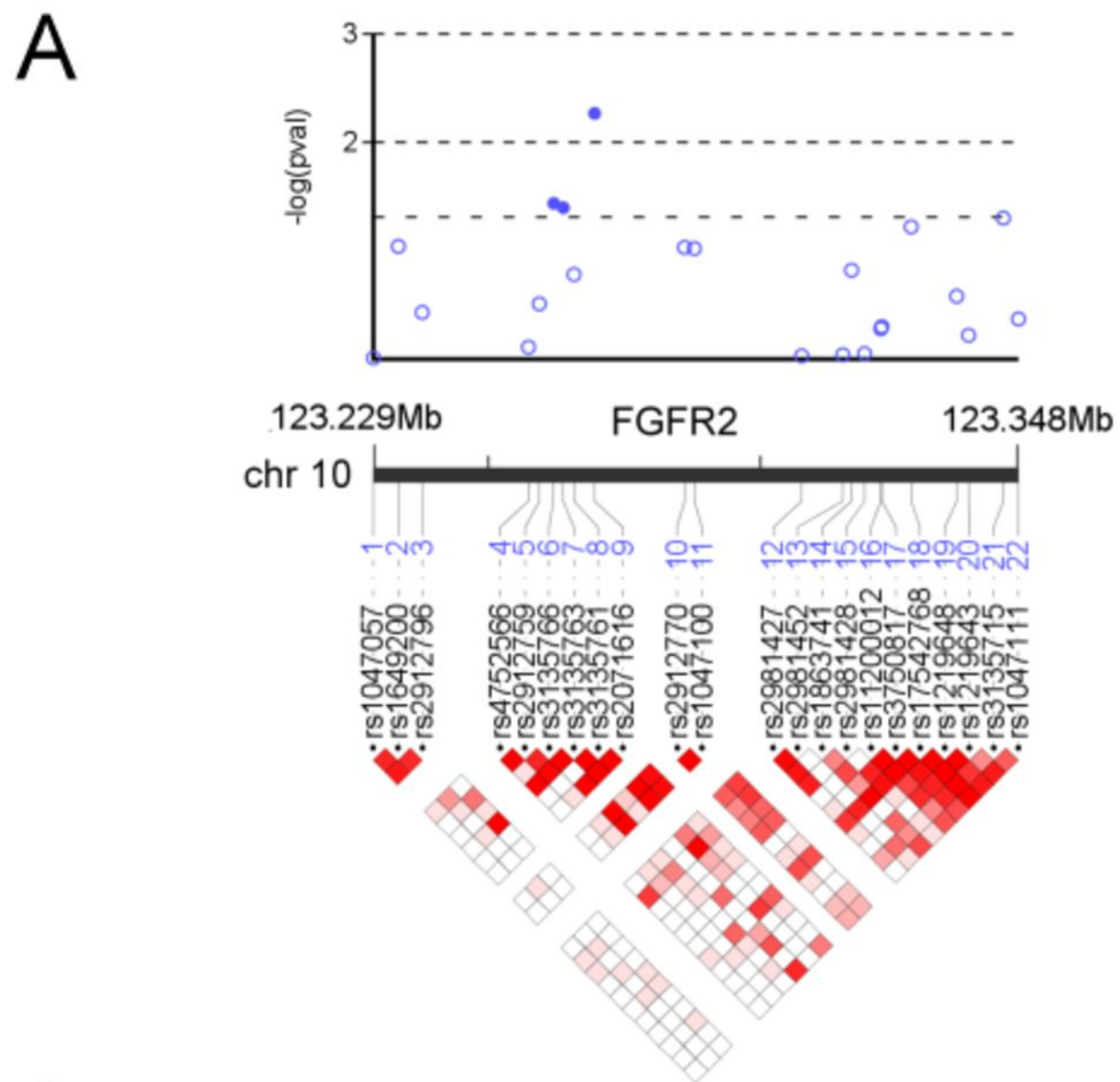
- 1 Moore DD: Physiology. Sister act. *Science* 2007; **316**: 1436-1438.
- 2 Kharitononkov A, Shiyanova TL, Koester A *et al*: FGF-21 as a novel metabolic regulator. *J Clin Invest* 2005; **115**: 1627-1635.
- 3 Inagaki T, Dutchak P, Zhao G *et al*: Endocrine regulation of the fasting response by PPARalpha-mediated induction of fibroblast growth factor 21. *Cell Metab* 2007; **5**: 415-425.
- 4 Kurosu H, Choi M, Ogawa Y *et al*: Tissue-specific expression of betaKlotho and fibroblast growth factor (FGF) receptor isoforms determines metabolic activity of FGF19 and FGF21. *J Biol Chem* 2007; **282**: 26687-26695.
- 5 Ogawa Y, Kurosu H, Yamamoto M *et al*: BetaKlotho is required for metabolic activity of fibroblast growth factor 21. *Proc Natl Acad Sci U S A* 2007; **104**: 7432-7437.
- 6 Suzuki M, Uehara Y, Motomura-Matsuzaka K *et al*: betaKlotho is required for fibroblast growth factor (FGF) 21 signaling through FGF receptor (FGFR) 1c and FGFR3c. *Mol Endocrinol* 2008; **22**: 1006-1014.
- 7 Kharitononkov A, Dunbar JD, Bina HA *et al*: FGF-21/FGF-21 receptor interaction and activation is determined by betaKlotho. *J Cell Physiol* 2008; **215**: 1-7.
- 8 Tomaszewski M, Charchar FJ, Lynch MD *et al*: Fibroblast growth factor 1 gene and hypertension: from the quantitative trait locus to positional analysis. *Circulation* 2007; **116**: 1915-1924.
- 9 Firmann M, Mayor V, Vidal PM *et al*: The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* 2008; **8**: 6.

- 10 Tobin MD, Sheehan NA, Scurrah KJ, Burton PR: Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005; **24**: 2911-2935.
- 11 Broeckel U, Hengstenberg C, Mayer B *et al*: A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 2002; **30**: 210-214.
- 12 Tomaszewski M, Charchar FJ, Samani NJ: Association studies in current cardiovascular genetics - functional variants, tags or both? *J Hum Hypertens* 2007; **21**: 425-426.
- 13 Horvath S, Xu X, Laird NM: The family based association test method: strategies for studying general genotype--phenotype associations. *Eur J Hum Genet* 2001; **9**: 301-306.
- 14 Yan J: Yet another Package for Generalized Estimating Equations. *R-News* 2002; **2/3**: 12-14.
- 15 Cupples LA, Arruda HT, Benjamin EJ *et al*: The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC Med Genet* 2007; **8 Suppl 1**: S1.
- 16 Samani NJ, Erdmann J, Hall AS *et al*: Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; **357**: 443-453.
- 17 Erdmann J, Grosshennig A, Braund PS *et al*: New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet.* 2009; **41**: 280-282.
- 18 Storey JD, Tibshirani R: Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003; **100**: 9440-9445.

- 19 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263-265.
- 20 Tomaszewski M, Brain NJ, Charchar FJ *et al*: Essential hypertension and beta2-adrenergic receptor gene: linkage and association analysis. *Hypertension* 2002; **40**: 286-291.
- 21 Kharitononkov A, Wroblewski VJ, Koester A *et al*: The metabolic state of diabetic monkeys is regulated by fibroblast growth factor-21. *Endocrinology* 2007; **148**: 774-781.
- 22 Badman MK, Pissios P, Kennedy AR, Koukos G, Flier JS, Maratos-Flier E: Hepatic fibroblast growth factor 21 is regulated by PPARalpha and is a key mediator of hepatic lipid metabolism in ketotic states. *Cell Metab* 2007; **5**: 426-437.
- 23 Potthoff MJ, Inagaki T, Satapati S *et al*: FGF21 induces PGC-1alpha and regulates carbohydrate and fatty acid metabolism during the adaptive starvation response. *Proc Natl Acad Sci U S A* 2009; **106**: 10853-10858.
- 24 Christodoulides C, Dyson P, Sprecher D, Tsintzas K, Karpe F: Circulating fibroblast growth factor 21 is induced by peroxisome proliferator-activated receptor agonists but not ketosis in man. *J Clin Endocrinol Metab* 2009; **94**: 3594-3601.
- 25 Moyers JS, Shiyanova TL, Mehrbod F *et al*: Molecular determinants of FGF-21 activity-synergy and cross-talk with PPARgamma signaling. *J Cell Physiol* 2007; **210**: 1-6.

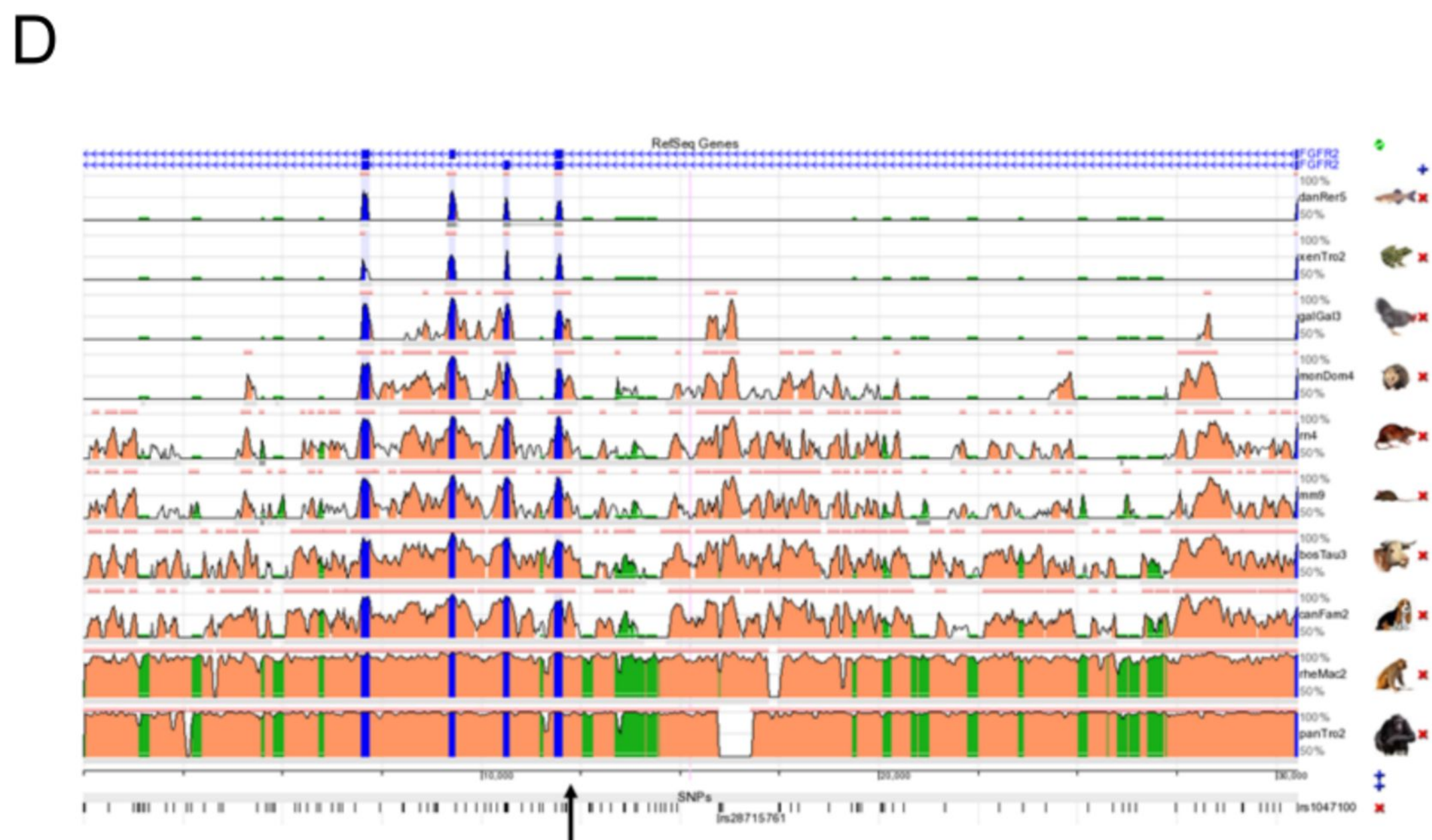
- 26 Hughes SE: Differential expression of the fibroblast growth factor receptor (FGFR) multigene family in normal human adult tissues. *J Histochem Cytochem* 1997; **45**: 1005-1019.
- 27 Yie J, Hecht R, Patel J *et al*: FGF21 N- and C-termini play different roles in receptor interaction and activation. *FEBS Lett* 2009; **583**: 19-24.
- 28 Katoh M: Cancer genomics and genetics of FGFR2 (Review). *Int J Oncol* 2008; **33**: 233-237.
- 29 de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008; **17**(R2): R122-128.





**C**

species	sequence
hedgehog	AAAA-GCCTCAAG-CTGCTGGCTGACACC----
pika	AAAA-GCCTCAAG-CTGCTGGCTGACAC-----
cow	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
<b>human</b>	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
chimp	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
orangutan	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
dog	-----CCTCAAG-CTGCTGGCTGACACCTTTG
mouse	-----CTCAAG-CTGCTGGCTGACACCTTTG
eur. hedgehog	-----AG-CTGCTGGCTGACACCTTTG
horse	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
dasypos	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
hyrax	-----CCTCAAG-CTGCTGGCTGACACCTTTG
gorilla	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG
lemur	AAAA-GCCTCAAG-CTGCTGGCTGACACCTTTG



1. **Case Study**

