Final report

During the project we carried out the following tasks and subprojects:

1. Collection of samples and clinical data; evaluation of clinical data
2. Utilization of the samples and evaluation of the results of our previous sublingual immunotherapy project (DesensIT)
3. Investigation of the role of the Hypo/YAP1 pathway in asthma
4. Investigation of the role of the Tie2 pathway in asthma and allergic conjunctivitis
5. Investigation of the genetic background and pathomechanism of asparaginase allergy
6. Investigation of plasma neutrophil extracellular trap levels in chronic respiratory diseases
7. Investigation of lncRNAs as circulating biomarkers in chronic respiratory diseases
8. Extracellular vesicles in asthma and allergic rhinitis.
9. Bioinformatic developments

Subprojects in details

1. Collection of samples and clinical data; evaluation of clinical data

Throughout the project we continuously collected different types of biological samples and clinical data from patients with the following diseases: Allergy (adult and children), asthma (adult and children), COPD (adult), acute lymphoblastic leukemia (children). We also collected samples and clinical data from corresponding controls. The number of the utilized samples can be seen in the different projects.

We also presented a request for ethical permission to ETT involving sample collection for the extracellular vesicle (EV) study and for collection of induced sputum. With the help of the participant physicians we worked out improved versions of questionnaires involving now additional questions related to the EV study. We worked out the protocol for sample processing for investigating the role of EVs in asthma.

We evaluated with conventional statistical method and with a Bayesian statistical framework the clinical data collected from severe asthma patients in Hungary. Our results were the following:

Data of 416 patients from the public pulmonary dispensaries were evaluated (group 1) and it was compared with data of 104 severe asthma patients (group 2) registered in the asthma outpatient clinic of National Korányi Institute of Pulmonology (NKI) (group 2) which served as a methodological center.

There was no difference between the groups in gender distribution, prevalence of allergy, systemic corticosteroid dependence, corticosteroid burst treatment and the mean value of personal best FEV1. Although the mean worst FEV1 values were statistically different, the difference was clinically not relevant (38.7% vs 41.7% in group 1 vs 2, respectively)

On the other hand, there were significant differences between the two groups in many respects. Average age of the patients was lower in group 1, while the duration of asthma was shorter in the dispensary group ($p = 0.0002$ and 0.01, respectively). Significantly more patients were in group 2, who developed asthma in childhood (age <12; 8.9% vs. 26.9% $p<0.0001$), and there were significantly more patients in group 2 with salicylate/NSAID intolerance (17.0% vs 39.4%; $p<0.0001$) and rhinosinusitis (32.9% vs. 84.4%; $p<0.0001$).

Analyzing the functional data of the entire study population showed that 72.1% of severe asthma subjects had persistent airflow limitation, defined as FEV1<80% (mean ± SD: 58.1±13.6% /of predictive), and in this regard there was no difference between the groups (72.1% for both groups). Severe airway obstruction, defined as FEV1≤60 % (of predictive), was in 37.9% of all severe asthmatics, with a mean FEV1 value of 47.5±9.5 %. In the remainder 62.1% of all patients the personal best FEV1 was above 60% (mean 79.3±13.8 %). There was no correlation between the personal best FEV1 and asthma duration.

Among patients where the disease started in childhood there were significantly more allergic than among patients with adult-onset asthma (80.9% vs 53.4%; p<0.0001, in childhood-onset vs adult-onset, respectively).

There was no difference in the proportion of patients with severe airway obstruction (best FEV1<60) between allergic and non-allergic severe asthmatics (34.1% vs 32.3%, respectively, ns). We could not find any relationship between systemic steroid dependence and non-allergic severe asthma as the ratio of allergic/non-allergic disease was similar in cases requiring maintenance systemic steroid treatment and in cases without that (100 allergic/92 non-allergic vs 185 allergic/125 non-allergic, 58.0/42.0% vs 55.9/44.1%, ns, with and without systemic steroid treatment, respectively).

For the calculation of direct causal relevance between patient characteristics, a Bayesian statistical framework was used, named Bayesian network based Bayesian multilevel analysis of relevance (BN-BMLA). Altogether 10 characteristics collected through the questionnaires were involved in this analysis and all the 520 patients were included. The calculated Bayesian dependency network can be seen in the on line paper in Figure 2. The network is an undirected graph where an edge between two nodes (here characteristics) represents direct casual relevance (i.e. a node is a direct cause of the other node), and its width is proportional to the probability of the corresponding nodes being directly relevant to each other.

This analysis confirmed some evident or previously shown dependencies, e.g. there was a strong connection between allergy and rhinosinusitis with an a posteriori value of (P=0.86), between disease onset (childhood or adult) and allergy (P=0.9), between best and worst FEV1 and also showed some novel ones.

E.g. there was a strong direct positive casual relevance between salicylate intolerance and rhinosinusitis (P=1), best FEV1 value and rhinosinusitis (P=1) and interestingly between best FEV1 and steroid burst therapy (P=0.91). This latter means that within the SA population the chance of oral steroid burst therapy increased with better lung functions when best FEV1 were considered. The connection was inverse in case of worst FEV1 and oral steroid burst therapy (P=0.92). There was a suggestive relevance between gender and salicylate intolerance (P=0.63). In this case 'M-' in the Figure 1 on the edge between the two nodes means that the prevalence of this characteristics was lower in males. There was also a suggestive relevance between oral systemic steroid treatment and salicylate intolerance (P=0.71), and age and oral steroid burst therapy (P=0.73).

The relevance revealed by BN-BMLA were confirmed with conventional statistics. All the above-mentioned connections proved to be statistically significant. E.g. SA patients with rhinosinusitis had on average higher best FEV1 than patients without this comorbidity (72.4±20.0% vs. 63.2±18.8%; p<0.0001). The mean age of patients who received corticosteroid burst therapy was younger (55.8±13.0 vs. 59.0±14.9 years; p=0.03). Significantly more patients received regular oral systemic corticosteroid treatment who had salicylate intolerance (44.1% vs. 30.8%; p=0.009), the proportion of rhinosinusitis was significantly higher in patients with salicylate intolerance (72.0% vs. 34.7%; p<0.0001; the proportion of rhinosinusitis with and without salicylate intolerance, respectively), the proportion of salicylate intolerance was significantly higher in patients with rhinosinusitis (36.2% vs. 10.3%; p<0.0001) and the proportion of salicylate intolerance was significantly higher in females (25.3% vs. 15.1%; p=0.007; proportion of salicylate intolerance in females and males, respectively). The group1 and group2 did not differ in these respects.

From the above mentioned results we published the following paper: Csoma Z, Gál Z, Gézsi A, Herjavecz I, Szalai C. Prevalence and characterization of severe asthma in Hungary. Sci Rep. 2020;10(1):9274.

In 2015 the PI of the project participated in Pediatric Allergy and Asthma Meeting in Berlin. During the meeting the Hungarian asthma research group joined to the Pharmacogenomics in Childhood Asthma (PiCA) consortium. The consortium was initiated to perform large-scale pharmacogenomics studies. In total, 14,016 children/young adults (up to 25 years) from 11 different countries are enrolled in the PiCA consortium. The children within the PiCA consortium are a good reflection of the global heterogeneous pediatric asthma population. Different outcome measures reflect different dimensions of asthma. Therefore, by classifying three commonly used outcome measures using data that are widely available within the PiCA consortium, we will be able to study distinct response phenotypes.
A published paper: Farzan N, … Szalai C, et al. Rationale and design of the multiethnic Pharmacogenomics in Childhood Asthma consortium. Pharmacogenomics. 2017 Jul;18(10):931-943

2. Utilization of the samples and evaluation of the results of our previous sublingual immunotherapy project (DesensIT)

We published a paper from the project:
Molnár V, Nagy A, Tamási L, Gálffy G, Böcskei R, Bikov A, Czaller I, Csoma Z, Krasznai M, Csáki C, Zsigmond G, Csontos Z, Kurucz A, Kurucz E, Fábos B, Bálint BL, Sasvári-Székely M, Székely A, Kótyuk E, Kozma GT, Cserta G, Farkas A, Gál Z, Gézsi A, Millinghoffer A, Antal P, Szalai C. From genomes to diaries: a 3-year prospective, real-life study of ragweed-specific sublingual immunotherapy. Immunotherapy. 2017 Nov;9(15):1279-1294.

The results and samples of the project are utilized in subprojects 4, 6 and 7. The evaluations of the results of the project are still in progress.

3. Investigation of the role of the Hypo/YAP1 pathway in asthma

The main aim of the study was to investigate the members of the Hippo pathway and compare their gene expression in the induced sputum of asthmatic patients and healthy controls. Moreover, it was also studied whether genetic variations in the YAP1 gene could influence the susceptibility of asthma or subgroups of asthma.
The gene expression analysis was done using the induced sputum of 20 asthmatic patients and 12 healthy controls. The genotyping analysis included 1235 unrelated individuals, out of which 525 were asthmatic children and 710 healthy controls. Real-time quantitative PCR was performed on *LATS1*, *LATS2*, *MST1*, *MST2*, *Ww45*, *YAP1*, *TAZ* and *β-actin* using an ABI 7900HT Fast Real-Time PCR System. *β-Actin* was used as an endogenous control and all results were normalized to it. Western blot analysis was carried out on human induced sputum samples. The KBiosciences

Competitive Allele-Specific PCR (KASP) version 4.0 genotyping system was used along with the TaqMan ABI 7900HT Fast Real-Time PCR system to genotype fourteen SNPs on the *YAP1* gene. For the evaluation we used traditional statistical methods (like Mann–Whitney *U* test, Kruskal–Wallis test, Fisher's exact test, Spearman non-parametric test) as well as BN-BMLA developed by our research group.

In the induced sputum of 18 asthma patients and 10 control subjects we measured the gene expression level of 7 members of the Hippo/YAP1 pathway. The expression of all genes could be detected in both cases and controls. The mean gene expression level of *YAP1* was slightly lower in asthmatic than in control patients.

During the correlation studies we found a significant and positive correlation between *YAP1* mRNA level and the sputum bronchial epithelial cells (r=0.575, p=0.003). There was a significant and negative correlation between *TAZ* mRNA and sputum neutrophils (r = -0.509, p=0.009) and *STK4* showed a significant and positive correlation with sputum eosinophils (r=0.425, p=0.034)..

We examined whether any of the SNPs in the *YAP1* gene influence the susceptibility of asthma or the different phenotypes. The genotyping analysis included 1233 unrelated individuals, out of which 522 were asthmatic children and 711 healthy controls. There was no significant association with any of the SNPs and asthma susceptibility, allergic status, inhalative, outdoor, indoor allergies, allergic and non-allergic asthma, comorbidities of rhinitis and conjunctivitis or serum IgE and eosinophil levels. However, SNP rs2846836 was significantly associated with exercise-induced asthma (OR=2.1 [1.3-3.4], p=0.004, power=0.83). Additionally, distribution of genotypes of SNP rs11225138 showed a significant difference between GINA 1-2 and GINA 3-4 statuses in a dominant model (OR=2.8 [1.4-5.6], p=0.003, power=0.83).

In order to find more evidence for the associations, we also conducted haplotype analyses. We found a significant difference between patients of GINA 2 and GINA 3 when we compared the frequencies of a haplotype formed by the rare alleles of SNPs rs1426398 and rs11225138, where the frequency of TC haplotype was more prevalent in GINA 3 than in GINA 2 (28% vs. 8%; $p=10^{-7}$). Furthermore, the CA haplotype from SNPs rs11225138 and rs1426394, also showed a significant difference when patients with GINA 3 were compared to GINA 2 asthmatics (26% vs. 7%, $p=10^{-7}$). When we included more than two SNPs in the analysis, additional associations were found.

Western blots were carried out on those proteins of the Hippo/YAP1 pathway whose genetic variations showed significant associations with asthma or phenotypes.

The signal for the FRMD6 protein could be detected in all sputum samples from both asthmatic and control patients. Interestingly, however, the YAP1 protein could not be detected in the sputum samples of the healthy controls, it was well-seen in the sputum samples of the mild asthmatics (GINA 1,2) and was also absent from the sputum of severe (GINA 3,4) asthmatics.

Based on the genotyping results of 29 SNPs in *YAP1*, *FRMD6* and *BIRC5* genes, laboratory data and characteristics of the asthmatic participants, the a posteriori probabilities of relevance between the variables with respect to target variables were calculated by BN-BMLA.

As expected, e.g. IgE levels or inhalative allergy are highly relevant to allergic asthma or, eosinophil levels and allergic conjunctivitis to allergic rhinitis.

In the case of genetic variations, no direct SNP-SNP or gene-gene interactions were found. The most relevant association was between rs9671722 in the *FRMD6* gene and exercise-induced asthma with a posterior probability of strong relevance of 0.99. The network structure suggested a direct relevance of rs9671722 to exercise-induced asthma, while another SNP (rs3751464) of the *FRMD6* gene was found to be directly relevant to allergic rhinitis and transitively associated through allergic rhinitis with exercise-induced asthma.

The paper about the above described results was published in Allergy, Asthma & Immunology Research journal: Fodor LE, Gézsi A, Ungvári I, Semsei ÁF, Gál Z, Nagy A, Gálffy G, Tamási L, Kiss A, Antal P, Szalai C. Investigation of the possible role of the Hippo/YAP1 pathway in asthma and allergy. Allergy Asthma Immunol Res. 2017 May;9(3):247-256.

4. Investigation of the role of the Tie2 pathway in asthma and allergic conjunctivitis

In this subproject two consecutive studies were carried out. First we investigated the role of three SNPs in the TEK gene asthma and allergic diseases, then with an expanded population (involving patients from the DesensIT project) we screened the TEK gene with 112 SNPs for association with conjunctivitis and asthma, then the selected SNPs were studied in another population. Here is the summaries of the two projects:

1. The Tie2 receptor is an important player in angiogenesis. The Tie2 mRNA and protein are abundantly expressed in the lungs and the associated pathway also has an important role in the development and function of the eye. Tie2 is encoded by the TEK gene in humans. Recently, variations in the TEK gene have been found associated with asthma. To investigate whether variations in the TEK gene influenced the susceptibility to pediatric asthma and/or associated phenotypes like GINA status, viral- or exercise-induced asthma, allergic asthma, indoor, outdoor, inhalative allergies, IgE and eosonophil levels, allergic rhinitis and allergic conjunctivitis. Three single nucleotide polymorphisms (SNPs, rs3780315, rs581724 and rs7876024) in the TEK gene were genotyped in 1189 unrelated individuals, out of which 435 were asthmatic children and 754 healthy controls. Different types of asthma, allergies and co-morbidities were defined in 320 patients. Among the fully phenotyped 320 asthmatic patients 178 (55.6%) also had allergic rhinitis and 100 (31.3%) had conjunctivitis. Among the rhinitis patients 98 (55.1%) also had conjunctivitis. Two patients had conjunctivitis without rhinitis. The genotyped SNPs showed no association with asthma. However, SNP rs581724 was significantly associated with allergic conjunctivitis in a recessive way (p=0.007; OR=2.3 (1.3-4.4)) within the asthmatic population. The risk remained significant when the whole population (asthmatics and healthy controls) was included in the calculation (p = 0.003; OR = 2.1 (1.3-3.6)). The minor allele of the rs581724 SNP which is associated with the increased risk to conjunctivitis is also associated with reduced Tie2 expression. CONCLUSIONS: There was a significant association between SNP rs581724 and the occurrence of allergic conjunctivitis in asthmatic children. If additional studies can confirm the role of the Tie2 pathway in allergic conjunctivitis, it can be a potential novel therapeutic target in the disease.

2. Tie2, coded by the TEK gene, is a tyrosine kinase receptor and plays a central role in vascular stability. It was suggested that variations in the TEK gene might influence the susceptibility to asthma and allergic conjunctivitis. The aim of this study was to further investigate these suggestions, involving different populations and to study the Tie2 related pathway on a mouse model of asthma. The discovery, stage I cohort involved 306 patients with moderate and severe allergic rhinitis, the stage II study consisted of 4 cohorts, namely adult and pediatric asthmatics and corresponding controls. Altogether, there were 1258 unrelated individuals in these cohorts, out of which 63.9% were children and 36.1% were adults. In stage I, 112 SNPs were screened in the TEK gene of the patients in order to search for associations with asthma and allergic conjunctivitis. The top associated SNPs were selected for association studies on the replication cohorts. The rs3824410 SNP was nominally associated with a reduced risk to of asthma in the stage I cohort and with severe asthma within the asthmatic population (p=0.009; OR=0.48) in the replication cohort. In the stage

I study, 5 SNPs were selected in conjunctivitis. Due to the low number of adult patients with conjunctivitis, only children were involved in stage II. Within the asthmatic children, the rs622232 SNP was associated with conjunctivitis in boys in the dominant model (p=0.004; OR=4.76), while the rs7034505 showed association to conjunctivitis in girls (p=0.012; OR=2.42). In the lung of a mouse model of asthma, expression changes of 10 Tie2 pathway- related genes were evaluated at three points in time. Eighty percent of the selected genes showed significant changes in their expressions at least at one time point during the process, leading from sensitization to allergic airway inflammation. The expressions of both the Tek gene and its ligands showed a reduced level at all time points. In conclusion, our results provide additional proofs that the Tie2 pathway, the TEK gene and its variations might have a role in asthma and allergic conjunctivitis. The gene and its associated pathways can be potential therapeutic targets in both diseases.

From the results of these two projects, two papers were published:
Fodor LE, Gézsi A, Gál Z, Nagy A, Kiss A, Bikov A, Szalai C. Variation in the TEK gene is not associated with asthma but with allergic conjunctivitis. Int J Immunogenet. 2018 Jun;45(3):102-108.
Gál Z, Gézsi A, Molnár V, Nagy A, Kiss A, Sultész M, Csoma Z, Tamási L, Gálffy G, Bálint BL, Póliska S, Szalai C. Investigation of the Possible Role of Tie2 Pathway and TEK Gene in Asthma and Allergic Conjunctivitis. Front Genet. 2020;11:128.

5. Investigation of the genetic background and pathomechanism of asparaginase allergy

Asparaginase is a pivotal component of pediatric acute lymphoblastic leukemia (ALL) treatment. Unfortunately, hypersensitivity can occur frequently against the enzyme which can lead to lower exposure to asparaginase resulting in suboptimal treatment response. In addition, hypersensitivity reactions to asparaginase ranges in severity from mild, transient (flushing or rash) to generalized anaphylaxis, which can be a potential threat to life. The primary goal of our study was to test the associations of HLA class II alleles with E. coli asparaginase hypersensitivity in a Hungarian population of 359 pediatric ALL patients by using next-generation sequence based typing of HLA-DRB1 and HLA-DQB1 alleles. Among 359 patients, the high-resolution sequence-based typing resulted in 35 unique HLA-DRB1 and 19 unique HLA-DQB1 alleles. Applying Bonferroni correction ($p \leq 2.66 \times 10^{-4}$) in multivariate logistic regression analyses HLA-DRB1*07:01 and HLA-DQB1*02:02 alleles showed significant associations with E. coli asparaginase hypersensitivity. In order to further investigate the relationship between HLA-DRB1*07:01 and HLA-DQB1*02:02 alleles associated with asparaginase hypersensitivity, we estimated haplotypes using the PHASE software. Two haplotypes containing HLA-DRB1*07:01 allele were estimated among patients: HLA-DRB1*07:01–HLA-DQB1*02:02 and HLA-DRB1*07:01–HLA-DQB1*03:03. Out of these, only HLA-DRB1*07:01–HLA-DQB1*02:02 showed positive association with asparaginase hypersensitivity. Next, the three-gene haplotypes were independently inferred to impute HLA-DQA1 alleles by using reference data from The Allele Frequency Net Database. Multivariate logistic regression analysis showed that HLA-DQA1*02:01 allele and HLA-DRB1*07:01–HLA-DQA1*02:01–HLA-DQB1*02:02 haplotype were positively and significantly associated with asparaginase hypersensitivity.
Next we investigated a possible cost-effective genetic testing method to identify patients harboring the risk HLA haplotype, creating the opportunity for a safer asparaginase treatment in their case.
We selected four HLA-tagging single nucleotide polymorphisms (SNPs) and determined them in 241 Hungarian patients with pediatric ALL. Based on the previous study, the HLA-DRB1 and

HLA-DQB1 alleles of the patients were known, so we were able to evaluate the performance of the SNPs on tagging for the HLA-DRB1*07:01-DQA1*02:01-DQB1*02:02 haplotype. The clinical utility of the markers were also evaluated. We identified a combination of two SNPs, rs28383172 and rs7775228 as a tags for HLA-DRB1*07:01-DQA1*02:01-DQB1*02:02 haplotype with a sensitivity and specificity values greater than 95%. Applying these two markers, the development of E. coli ASP hypersensitivity was predicted with specificity and negative predictive values of 88.1% and 65.6%, respectively. Compared to the rest of the population, patients with hypersensitivity-prone genotype would benefit more from the administration of less immunogenic pegylated asparaginase (PEG ASP) before the hypersensitivity evolves. Conclusions: The combination of rs28383172 and rs7775228 is suitable for identifying HLA-DRB1*07:01-DQA1*02:01-DQB1*02:02 haplotype carriers. A genotype-based drug choice would require very little extra cost compared to a strategy with PEG ASP therapy as frontline treatment for all. Further prospective studies are required to support the clinical utility of the method.

From this study a paper was published: Kutszegi N, Yang X, Gézsi A, Schermann G, Erdélyi DJ, Semsei ÁF, Gábor KM, Sági JC, Kovács GT, Falus A, Zhang H, Szalai C. HLA-DRB1*07:01-HLA-DQA1*02:01-HLA-DQB1*02:02 haplotype is associated with a high risk of asparaginase hypersensitivity in acute lymphoblastic leukemia. Haematologica. 2017 Sep;102(9):1578-1586.
And:
Kutszegi N, Semsei ÁF, Gézsi A, Sági JC, Nagy V, Csordás K, Jakab Z, Lautner-Csorba O, Gábor KM, Kovács GT, Erdélyi DJ, Szalai C. Subgroups of Paediatric Acute Lymphoblastic Leukaemia Might Differ Significantly in Genetic Predisposition to Asparaginase Hypersensitivity. PLoS One. 2015;10(10):e0140136.

From the results regarding the tag SNP an additional paper has already been submitted for publication.

6. Investigation of plasma neutrophil extracellular trap levels in chronic respiratory diseases

A flow cytometry-based method was developed to quantify in vivo circulating neutrophil extracellular trap (NET) levels in plasma and compare them in patients with different chronic inflammatory lung diseases. Seventeen asthmatic and 11 control children, 12 adult controls, 46 asthmatic, 6 COPD and 6 adult patients with asthma-COPD overlap syndrome (ACOS) were recruited in the study. The presence of NETs in unstimulated cell-free plasma was confirmed and visualized by confocal laser-scanning microscopy. No significant differences were found in plasma NET levels between children and adults, children with or without asthma and adults with or without asthma, COPD or ACOS. When asthmatic patients were stratified according to their disease severity the average plasma NET level was significantly higher in asthmatic patients with more serious symptoms (adjusted p=0.027). Patients with poorer pulmonary functions had higher plasma NET levels which negatively correlated with the FEV1 values (r = -0.39, p=0.002). Patients who were medicated daily with inhaled corticosteroids (ICS) had significantly lower average plasma NET level than patients who did not or just occasionally used ICS (p=0.027). If further studies confirm the NET-lowering effect of ICS in the circulation, it can be utilized in diseases where NETosis contributes to the pathogenesis.

Published paper: Gál Z, Gézsi A, Pállinger É, Visnovitz T, Nagy A, Kiss A, Sultész M, Csoma Z, Tamási L, Gálffy G, Szalai C. Plasma neutrophil extracellular trap level is modified by disease severity and inhaled corticosteroids in chronic inflammatory lung diseases. Sci Rep. 2020;10(1):4320.

7. Investigation of lncRNAs as circulating biomarkers in chronic respiratory diseases

Numerous researches have confirmed that 70-90% of the human genome is transcribed into RNA but only 1.2 % have protein coding ability. Long non-coding RNAs (lncRNAs) are greater than 200 bp in length, building a major part of non-coding RNAs, but in the meantime the least characterized. In this study, first we measured the gene expression levels of 84 inflammatory response and autoimmunity associated lncRNAs in the blood of patients with mild or moderate (Global Initiative for Asthma (GINA) 1-3) and severe (GINA 4-5) asthma, COPD, and control patients (discovery cohort). Then, based on these results and the scientific literature we selected 6 lncRNAs and compared their expressions in an expanded population of patients with different chronic respiratory diseases including pediatric and adult asthma, mild and severe asthma, COPD, and in corresponding healthy controls. We also compared the expressions of these lncRNAs in different subgroups of asthma and investigated whether they could be used as biomarkers in these diseases.

Our research consisted of two stages. In the discovery cohort, 24 adult patients were involved, out of which 6 had mild or moderate asthma (GINA 1-3), 6 severe asthma (GINA 4-5), 6 COPD and 6 were healthy controls. The replication cohort consisted of 163 subjects. This cohort included 11 asthmatic children from the Allergology Department of Heim Pál Children's Hospital, 95 adult patients with asthma, 9 with COPD from the Asthma ambulance of National Korányi Institute of TB and Pulmonology, and from the Department of Pulmonology of Semmelweis University. Out of the asthmatic patients 31 had severe asthma (GINA 4-5) and 64 mild or moderate asthma (GINA 1-3).

In stage I the mean expression of 28 lncRNAs showed nominally significant differences ($P < 0.05$) in at least one comparison. Most differences were found between mild and severe asthma groups. In this comparison 23 out of 84 lncRNAs showed nominally significant differences. Nine lncRNAs showed expression differences between COPD and severe asthma, 3 between asthma and COPD, 3 between asthma and control, 9 between severe asthma and control, 1 between COPD and control, 2 between mild asthma and control groups.

In previous studies two lncRNAs (*OIP5-AS1, HNRNPU*) have been indirectly associated with eosinophil asthm. In our measurements, *HNRNPU* showed increased expression in severe asthma compared with mild asthma, while *OIP5-AS1* showed increased expression in COPD compared to asthma. Based on these differences, and data from the scientific literature and databases 6 lncRNAs were selected for the replication cohort (*OIP5-AS1, HNRNPU, RP11-325K4.3, JPX, RP11-282O18.3, AC016629.8*, later renamed to *MZF1-AS1*).

In the replication cohort three lncRNAs (*HNRNPU, RP11-325K4.3, JPX*) expressed significantly higher in pediatric controls than in adult controls. Because of this, the results when the two age groups were merged (e.g. in case of asthma) were excluded from the evaluations.

The largest and most differences were found between adult allergic rhinitis and control patients. In these cases, the mean expression levels of all 6 lncRNAs differed between the two groups. In respect of allergy, *OIP5-AS1* seemed to be the most important, since its mean expression level was significantly higher in all cases, where allergy was involved. It was also higher in allergic patients without asthma than in allergic asthmatic patients. In respect of these lncRNAs, allergic rhinitis differed most significantly from any other phenotypes. In allergic rhinitis the mean expressions of five lncRNAs (*RP11-325K4.3, OIP5-AS1, JPX, HNRNPU, MZF-AS1*) were significantly higher than in COPD, three (*OIP5-AS1, HNRNPU, JPX*) than in asthma, five (*OIP5-AS1, HNRNPU, RP11-325K4.3, RP11-282O18.3, JPX*) than in non-allergic asthma and one (*OIP5-AS1*) than in allergic asthma. Adult allergic and non-allergic asthma differed in the expression of three lncRNAs from each other, *RP11-325K4.3, HNRNPU* and *OIP5-AS1* expressed higher in allergic asthma. COPD and asthma differed in their expression of one lncRNA from each other. *RP11-325K4.3* expressed significantly higher in the blood of asthmatics than in patients with COPD.

In contrast to the discovery cohort, in this expanded population none of the lncRNAs showed association with asthma severity. No differences were found between pediatric asthma and controls. In the replication cohort similarly to the discovery cohort, *JPX* did not show a gender specific expression.

Next, we also analyzed whether the expressions of these lncRNAs differ in different subgroups of asthma. No differences were found when asthmatic patients were stratified according to their lung functions (FEV1< 80% vs. FEV1>80%), inhaled corticosteroid usage (regular vs. non-regular), severity, and controllability (controlled vs. non-controlled). We also tested whether the expression levels of these lncRNAs correlated with the blood eosinophil or neutrophil levels but found no correlation.

Next, we investigated, whether these lncRNAs can be used as diagnostic biomarkers for any studied chronic respiratory disease. For six relevant comparisons we created several Naïve Bayesian classifiers based on the normalized expression levels of different lncRNA combinations, namely (1) using each lncRNA alone, (2) all six lncRNAs (i.e. the full model), and (3) only those that showed statistically significant expression differences in case of the given comparison (i.e. the reduced model). Then, we assessed the performance of the classification models by computing their weighted accuracy (WA) utilizing a leave-one-out cross-validation scheme (see Methods).

Classifying adult allergic rhinitis patients and adult controls, three models achieved a very high performance (WA = 0.98 in case of (1) using *OIP5-AS1* alone, (2) using all six lncRNAs, which is the same model as (3) using all significant lncRNAs with respect to the given comparison). Clearly, these models utilized the high discriminative power of *OIP5-AS1*. Comparing adult COPD and adult patients with allergic rhinitis, using all five significant lncRNAs also resulted in a high performance (WA = 0.85).

In certain cases, combining all six lncRNAs resulted in significantly higher performance than any individual lncRNAs. Comparing adult allergic rhinitis and asthmatic patients, the best model using individual lncRNAs resulted in a WA of 0.53, however, combining all six lncRNAs resulted in a WA of 0.7. Similarly, comparing adult COPD and adult asthmatic patients, the best individual model had a WA of 0.53, and the full model had 0.61, respectively.

In other cases, using the combination of those lncRNAs that showed statistically significant expression differences resulted in a slightly higher performance than the full model. Namely, in case of the aforementioned comparison of adult COPD and adult allergic rhinitis patients, and in case of comparing adult allergic asthmatic and non-allergic asthmatic patients (WA = 0.65, and 0.68 in case of the full model, and the reduced model, respectively).

The *OIP5-AS1* lncRNA had the highest discriminative power in case of three out of the six comparisons. Moreover, comparing adult patients with allergic and adult non-allergic asthmatic patients, the model using the individual *OIP5-AS1* had the highest performance of all models (WA = 0.74, which is 5 percent point higher than the second-best model).

From the results of this study a paper has been submitted for publication.

8. Extracellular vesicles in asthma and allergic rhinitis.

Extracellular vesicles (EVs) are small, lipid membrane enclosed subcellular structures carrying biomolecules which are released by cells into their environment. EVs are not merely “innocent bystanders” but play important roles in physiology and pathology. In physiology, EVs contribute to homeostasis and promote host-defense mechanisms including haemostasis and inflammation, whereas in pathology EVs may contribute to disease development and progression. Because EVs are capable of delivering their proteins, lipid and RNA cargo to target (recipient) cells, EVs can regulate gene expression and consequently the phenotype and biological functions of the target cell. Thus, by exchanging information between cells, EVs contribute to intercellular communication. In this study we investigated, whether the protein content of plasma-derived EVs are different in asthmatic patients and in different asthma endotypes.

We isolated EVs from 6 patients with mild asthma, 6 from severe asthma and 6 from healthy controls. The protein content of the EVs was determined by mass spectrometry. In the EVs 142 proteins were identified. In the protein content significant differences were found between asthma and controls (8 proteins), severe vs mild asthma (8 proteins), according to medication usage (12 proteins), controlled asthma vs. non-controlled asthma (7 proteins) and allergic asthma vs. non-allergic asthma (4 proteins). The evaluation of these data is still in progress.

In connection to this project we also carried out a bioinformatic development and an analysis. We integrated publically and privately available genomic data sources into the Bayesian framework. For these purposes, we made several improvements to the newly developed Quantitative Semantic Fusion (QSF) System to assist genomic information fusion and a priori knowledge extraction.

The QSF System is an extensible framework that incorporates distinct annotated semantic types (also called: entities) and links between them by integrating different data sources from the Linked Open Data world. The QSF System then enables the users to quantitatively prioritize a freely chosen entity based on evidences propagated from any other, possibly multiple entities through the connecting links. Currently the system contains genes, taxa, diseases, phenotypes, disease categories (UMLS semantic types and MeSH disease classes), pathways, substances, assays, cell lines and the targets of the compounds. Links define associations between entities. For example, genes and pathways are connected with a link which represent gene-pathway associations.

To demonstrate the utility of the developed methodology, we used the QSF System to quantitatively prioritize diseases and phenotypes that are associated with five, partly overlapping sets of genes known to be involved in the biogenesis and/or secretion of different types of extracellular vesicles (EVs). Furthermore, we downloaded and reanalyzed many publicly available gene expression experiments from the Gene Expression Omnibus that represent various diseases and disease conditions including asthma, acute lymphoblastic leukemia, sepsis and different types of cancer. Then, we computed the enrichment of the aforementioned EV gene lists using the Ensemble of Gene Set Enrichment Analysis method. We found that genes reported to participate in the biogenesis or secretion of EVs, are significantly associated with numerous common diseases, including different types of tumors and cardiovascular diseases.

The results were published: Gézsi A et al. Systems biology approaches to investigating the roles of extracellular vesicles in human diseases. Exp Mol Med. 2019;51(3):1-11.

9. Bioinformatic developments

Our first goal was to supplement the Bayesian network based Bayesian multilevel analysis of relevance (BN-BMLA) methodology to handle hybrid models, i.e. Bayesian networks containing both continuous and discrete variables). With these hybrid models, we can jointly analyze the effect of genetic polymorphisms and gene expression data on a given target variable (e.g. asthma status).

As a preliminary goal, first we concentrated on the problem of accurate prediction of genetic variants from next-generation sequencing (NGS) results. The reason behind this was the fact that the level of uncertainty in NGS measurements is still higher than expected, i.e. currently there is no single best general individual variant calling method with both superior sensitivity and precision and there are significant discrepancies between commonly used variant calling pipelines. However, as variant callers produce a rich set of annotations that provide abundant information about variant quality and various biases, we hypothesized that this information could be combined into a better performing overall variant calling "ensemble" model. Therefore, we constructed a software, called VariantMetaCaller, which combines annotation information from various variant callers using Support Vector Machines. Our novel method predicts the probability that a variant is a true genetic variant and not a sequencing artefact, which provides a principled solution for quantitative support for variant filtering. Using artificially generated and real sequencing data, we demonstrated the usefulness of intermediate information fusion, by showing that VariantMetaCaller outperformed individual variant callers and a late information fusion method under a wide range of conditions. We published the results in a methodology article: Gézsi A, Bolgár B, Marx P, Sarkozy P, Szalai C, Antal P. VariantMetaCaller: automated fusion of variant calling pipelines for quantitative, precision-based filtering. BMC Genomics. 2015;16:875.

Next, to enable the measurement of the performance of various structure learning methods from combined genetic variation and gene expression data, we implemented a software, called GenomicDataGenerator which can be used to generate artificial hybrid genomic data. The software first constructs a randomly generated Bayesian model (called reference model) using the parametrization defined in a configuration file. The parameters specify the number of genes, the mean number of genetic variants per gene, the allele frequency spectrum of the genetic variants, the effect size distribution (mean and standard deviance) of the genetic variants on the gene expression levels, and the mean number and the effect size distribution of genes (i.e. transcription factors) affecting the expression of other genes. Then, the software generates data based on the reference model. The generated data can then be used to train various structure learning methods, and the performance of these methods can be measured with respect to the ability how well they reproduce the reference model.

Next, we investigated a specific class of hybrid dynamic Bayesian networks, called Hidden Markov Models (HMM) with Gaussians as emission probability distribution. These models can be efficiently used for analyzing the temporal characteristics of gene expression, i.e. to account for the horizontal dependencies along the time axis in time course expression data. Time course experiments are increasingly popular, as dynamic, temporal gene expression profiles provide an important characterization of gene function. We implemented several types and extensions of the HMMs, e.g. standard ergodic model (where every element of the transition probability matrix is positive) and left-right models (where only the elements of the upper diagonal part of the transition probability matrix are positive), and models with multiple Gaussian mixture components. With these models, the effect of a genetic polymorphism can be modeled as the mixture component which affects the observed probability distribution of gene expression. We evaluated and compared the performance of the various model types using synthetically generated (by GenomicDataGenerator) polymorphism and gene expression data sets. The assessment and the publication of the results are in progress.

The results of the development of the Quantitative Semantic Fusion System is described in the previous section.