Final report for KH-189581

The major goal of our study was to identify biomarkers of survival and therapy response in breast cancer an in other solid tumors. We have utilized bioinformatic analyses, genomic tests, cell culture methods and also investigated clinical samples.

As this is a KH project, which is not a thematic call, there are a two projects not directly related to the research topic, but I have also summarized these for the completeness of the overview of executed research projects.

Below I provide a short summary for each of the studies published. The two very last summaries are the most important publications related to the original grant application: in these papers we have summarized the online platform which enables the analysis of the developed database and pipeline for any researcher in the future.

Projects marked by * are those where I served either as corresponding/first/last author.

***Title: Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma**

Background: Potential prognostic biomarker candidates for hepatocellular carcinoma (HCC) are abundant, but their generalizability is unexplored. We cross-validated markers of overall survival (OS) and vascular invasion in independent datasets. Methods: The literature search yielded 318 genes related to survival and 52 related to vascular invasion. Validation was performed in three datasets (RNA-seq, n = 371; Affymetrix arrays, n = 91; Illumina gene chips, n = 135) by uni- and multivariate Cox regression and Mann–Whitney U-test, separately for Asian and Caucasian patients. Results: One hundred and eighty biomarkers remained significant in Asian and 128 in Caucasian subjects at $p < 0.05$. After multiple testing correction BIRC5 ($p = 1.9 \times 10-10$), CDC20 ($p = 2.5 \times 10-9$) and PLK1 ($p = 3 \times 10-9$) endured as best performing genes in Asian patients; however, none remained significant in the Caucasian cohort. In a multivariate analysis, significance was reached by stage ($p = 0.0018$) and expression of CENPH ($p = 0.0038$) and CDK4 ($p = 0.038$). KIF18A was the only gene predicting vascular invasion in the Affymetrix and Illumina cohorts ($p = 0.003$ and $p = 0.025$, respectively). Conclusion: Overall, about half of biomarker candidates failed to retain prognostic value and none were better than stage predicting OS. Impact: Our results help to eliminate biomarkers with limited capability to predict OS and/or vascular invasion. This study was published in Royal Society Open Science (https://doi.org/10.1098/rsos.181006).

***Title: Elevated HOX gene expression in acute myeloid leukemia is associated with NPM1 mutations and poor survival**

Acute myeloid leukemia (AML) is a clonal disorder of hematopoietic progenitor cells and the most common malignant myeloid disorder in adults. Several gene mutations such as in NPM1 (nucleophosmin 1) are involved in the pathogenesis and progression of AML. The aim of this study was to identify genes whose expression is associated with driver mutations and survival outcome. Genotype data (somatic mutations) and gene expression data including RNA-seq, microarray, and qPCR data were used for the analysis. Multiple datasets were utilized as training sets (GSE6891, TCGA, and GSE1159). A new clinical sample cohort (Semmelweis set) was established for *in vitro* validation. Wilcoxon analysis was used to identify genes with expression alterations between the mutant and wild type samples. Cox regression analysis was performed to examine the association between gene expression and survival outcome. Data analysis was performed in the R statistical environment. Eighty-five genes were identified with significantly altered expression when comparing NPM1 mutant and wild type patient groups in the GSE6891 set. Additional training sets were used as a filter to condense the six most significant genes associated with NPM1 mutations. Then, the expression changes of these six genes were confirmed in the Semmelweis set: HOXA5 ($P = 3.06E{-}12$, FC = 8.3), HOXA10 ($P = 2.44E{-}09$, FC = 3.3), HOXB5 ($P = 1.86E{-}13$, FC = 37), MEIS1 ($P = 9.82E{-}10$, FC = 4.4), PBX3 ($P = 1.03E{-}13$, FC = 5.4) and ITM2A ($P = 0.004$, FC = 0.4). Cox regression analysis showed that higher expression of these genes – with the exception of ITM2A – was associated with worse overall survival. Higher expression of the HOX genes was identified in tumors harboring NPM1 gene mutations by computationally linking genotype and gene expression. *In vitro* validation of these genes supports their potential therapeutic application in AML. We have published this study in Journal of Advanced Reserch (https://doi.org/10.1016/j.jare.2019.05.006).

***Title: Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology**

Whole exome sequencing (WES) enables the analysis of all protein coding sequences in the human genome. This technology enables the investigation of cancer-related genetic aberrations that are predominantly located in the exonic regions. WES delivers high-throughput results at a reasonable price. Here, we review analysis tools enabling utilization of WES data in clinical and research settings. Technically, WES initially allows the detection of single nucleotide variants (SNVs) and copy number variations (CNVs), and data obtained through these methods can be combined and further utilized. Variant calling algorithms for SNVs range from standalone tools to machine learning-based combined pipelines. Tools for

CNV detection compare the number of reads aligned to a dedicated segment. Both SNVs and CNVs help to identify mutations resulting in pharmacologically druggable alterations. The identification of homologous recombination deficiency enables the use of PARP inhibitors. Determining microsatellite instability and tumor mutation burden helps to select patients eligible for immunotherapy. To pave the way for clinical applications, we have to recognize some limitations of WES, including its restricted ability to detect CNVs, low coverage compared to targeted sequencing, and the missing consensus regarding references and minimal application requirements. Recently, Galaxy became the leading platform in non-command line-based WES data processing. The maturation of next-generation sequencing is reinforced by Food and Drug Administration (FDA)-approved methods for cancer screening, detection, and follow-up. WES is on the verge of becoming an affordable and sufficiently evolved technology for everyday clinical use. We published this project in Cancers (https://doi.org/10.3390/cancers11111725).

Title: FAK activity sustains intrinsic and acquired ovarian cancer resistance to platinum chemotherapy

Gene copy number alterations, tumor cell stemness, and the development of platinum chemotherapy resistance contribute to high-grade serous ovarian cancer (HGSOC) recurrence. Stem phenotypes involving Wnt-β-catenin, aldehyde dehydrogenase activities, intrinsic platinum resistance, and tumorsphere formation are here associated with spontaneous gains in Kras, Myc and FAK (KMF) genes in a new aggressive murine model of ovarian cancer. Adhesion-independent FAK signaling sustained KMF and human tumorsphere proliferation as well as resistance to cisplatin cytotoxicity. Platinum-resistant tumorspheres can acquire a dependence on FAK for growth. Accordingly, increased FAK tyrosine phosphorylation was observed within HGSOC patient tumors surviving neo-adjuvant chemotherapy. Combining a FAK inhibitor with platinum overcame chemoresistance and triggered cell apoptosis. FAK transcriptomic analyses across knockout and reconstituted cells identified 135 targets, elevated in HGSOC, that were regulated by FAK activity and β-catenin including Myc, pluripotency and DNA repair genes. These studies reveal an oncogenic FAK signaling role supporting chemoresistance. We published the results of this study in eLIfe (https://doi.org/10.7554/eLife.47327.001).

Title: Low level of exosomal long non-coding RNA HOTTIP is a prognostic biomarker in colorectal cancer

Molecular risk stratification of colorectal cancer can improve patient outcome. A panel of lncRNAs (*H19, HOTTIP, HULC* and *MALAT1*) derived from serum exosomes of patients with non-metastatic CRC and healthy donors was analyzed. Exosomes from healthy donors

carried significantly more *H19, HULC* and *HOTTIP* transcripts in comparison to CRC patients. Correlation analysis between lncRNAs and clinical data revealed a statistical significance between low levels of exosomal *HOTTIP* and poor overall survival. This was confirmed by multivariate analysis that *HOTTIP* is an independent prognostic marker for overall survival (HR: 4.5, CI: 1.69–11.98, p = 0.0027). Here, *HOTTIP* poses to be a valid biomarker for patients with a CRC to predict post-surgical survival time. The results of this project were published in RNA Biology (https://doi.org/10.1080/15476286.2019.1637697).

***Title: Independent validation of induced overexpression efficiency across 242 experiments shows a success rate of 39%**

Although numerous studies containing induced gene expression have already been published, independent authentication of their results has not yet been performed. Here, we utilized available transcriptomic data to validate the achieved efficiency in overexpression studies. Microarray data of experiments containing cell lines with induced overexpression in one or more genes were analyzed. All together 342 studies were processed, these include 242 different genes overexpressed in 184 cell lines. The final database includes 4,755 treatment-control sample pairs. Successful gene induction (fold change induction over 1.44) was validated in 39.3% of all genes at $p < 0.05$. Number of repetitions within a study ($p < 0.0001$) and type of used vector ($p = 0.023$) had significant impact on successful overexpression efficacy. In summary, over 60% of studies failed to deliver a reproducible overexpression. To achieve higher efficiency, robust and strict study design with multi-level quality control will be necessary. We have published this analysis in Scientific Reports (https://doi.org/10.1038/s41598-018-36122-8).

Title: Identifying Cancers Impacted by CDK8/19

CDK8 and CDK19 Mediator kinases are transcriptional co-regulators implicated in several types of cancer. Small-molecule CDK8/19 inhibitors have recently entered or are entering clinical trials, starting with breast cancer and acute myeloid leukemia (AML). To identify other cancers where these novel drugs may provide benefit, we queried genomic and transcriptomic databases for potential impact of CDK8, CDK19, or their binding partner CCNC. sgRNA analysis of a panel of tumor cell lines showed that most tumor types represented in the panel, except for some central nervous system tumors, were not dependent on these genes. In contrast, analysis of clinical samples for alterations in these genes revealed a high frequency of gene amplification in two highly aggressive subtypes of prostate cancer and in some cancers of the GI tract, breast, bladder, and sarcomas. Analysis of survival correlations identified a group of cancers where CDK8 expression correlated with shorter survival (notably breast, prostate, cervical cancers, and esophageal

adenocarcinoma). In some cancers (AML, melanoma, ovarian, and others), such correlations were limited to samples with a below-median tumor mutation burden. These results suggest that Mediator kinases are especially important in cancers that are driven primarily by transcriptional rather than mutational changes and warrant an investigation of their role in additional cancer types. This study was published in Cells (https://doi.org/10.3390/cells8080821).

Title: Evaluating ZNF217 mRNA Expression Levels as a Predictor of Response to Endocrine Therapy in ER+ Breast Cancer

ZNF217 is a candidate oncogene with a wide variety of deleterious functions in breast cancer. Here, we aimed at investigating in a pilot prospective study the association between *ZNF217* mRNA expression levels and the clinical response to neoadjuvant endocrine therapy (ET) in postmenopausal ER-positive (ER+) breast cancer patients. Core surgical biopsy samples before treatment initiation and post-treatment were obtained from 68 patients, and Ki-67 values measured by immunohistochemistry (IHC) were used to identify responders (n = 59) and non-responders (n = 9) after 4 months of ET. We report for the first time that high *ZNF217* mRNA expression level measured by RT-qPCR in the initial tumor samples (pre-treatment) is associated with poor response to neoadjuvant ET. Indeed, the clinical positive response rate in patients with low *ZNF217* expression levels was significantly higher than that in those with high *ZNF217* expression levels (P = 0.027). Additionally, a retrospective analysis evaluating *ZNF217* expression levels in primary breast tumor of ER+/HER2-/LN0 breast cancer patients treated with adjuvant ET enabled the identification of poorer responders prone to earlier relapse (P = 0.013), while *ZNF217* did not retain any prognostic value in the ER+/HER2-/LN0 breast cancer patients who did not receive any treatment. Altogether, these data suggest that *ZNF217* expression might be predictive of clinical response to ET. We published this study in Frontiers in Pharmacology (https://doi.org/10.3389/fphar.2018.01581).

***Title: Principles of tumorigenesis and emerging molecular drivers of SHH-activated medulloblastomas**

SHH-activated medulloblastomas (SHH-MB) account for 25–30% of all medulloblastomas (MB) and occur with a bimodal age distribution, encompassing many infant and adult, but fewer childhood cases. Different age groups are characterized by distinct survival outcomes and age-specific alterations of regulatory pathways. Here, we review SHH-specific genetic aberrations and signaling pathways. Over 95% of SHH-MBs contain at least one driver event – the activating mutations frequently affect sonic hedgehog signaling (PTCH1, SMO, SUFU), genome maintenance (TP53), and chromatin modulation (KMT2D, KMT2C, HAT complexes),

while genes responsible for transcriptional regulation (MYCN) are recurrently amplified. SHH-MBs have the highest prevalence of damaging germline mutations among all MBs. TP53-mutant MBs are enriched among older children and have the worst prognosis among all SHH-MBs. Numerous genetic aberrations, including mutations of TERT, DDX3X, and the PI3K/AKT/mTOR pathway are almost exclusive to adult patients. We elaborate on the newest development within the evolution of molecular subclassification, and compare proposed risk categories across emerging classification systems. We discuss discoveries based on preclinical models and elaborate on the applicability of potential new therapies, including BET bromodomain inhibitors, statins, inhibitors of SMO, AURK, PLK, cMET, targeting stem-like cells, and emerging immunotherapeutic strategies. An enormous amount of data on the genetic background of SHH-MB have accumulated, nevertheless, subgroup affiliation does not provide reliable prediction about response to therapy. Emerging subtypes within SHH-MB offer more layered risk stratifications. Rational clinical trial designs with the incorporation of available molecular knowledge are inevitable. Improved collaboration across the scientific community will be imperative for therapeutic breakthroughs. This study was published in Annals of Cliical and Translational Neurology (https://doi.org/10.1002/acn3.762).

***Title: Molecular markers and potential therapeutic targets in non-WNT/non-SHH (group 3 and group 4) medulloblastomas**

Childhood medulloblastomas (MB) are heterogeneous and are divided into four molecular subgroups. The provisional non-wingless-activated (WNT)/non-sonic hedgehog-activated (SHH) category combining group 3 and group 4 represents over two thirds of all MBs, coupled with the highest rates of metastases and least understood pathology. The molecular era expanded our knowledge about molecular aberrations involved in MB tumorigenesis, and here, we review processes leading to non-WNT/non-SHH MB formations.The heterogeneous group 3 and group 4 MBs frequently harbor rare individual genetic alterations, yet the emerging profiles suggest that infrequent events converge on common, potentially targetable signaling pathways. A mutual theme is the altered epigenetic regulation, and in vitro approaches targeting epigenetic machinery are promising. Growing evidence indicates the presence of an intermediate, mixed signature group along group 3 and group 4, and future clarifications are imperative for concordant classification, as misidentifying patient samples has serious implications for therapy and clinical trials.To subdue the high MB mortality, we need to discern mechanisms of disease spread and recurrence. Current preclinical models do not represent the full scale of group 3 and group 4 heterogeneity: all of existing group 3 cell lines are MYC-amplified and most mouse models resemble MYC-activated MBs. Clinical

samples provide a wealth of information about the genetic divergence between primary tumors and metastatic clones, but recurrent MBs are rarely resected. Molecularly stratified treatment options are limited, and targeted therapies are still in preclinical development. Attacking these aggressive tumors at multiple frontiers will be needed to improve stagnant survival rates. The paper for this project appreared in Journal of Hematology & Oncology (https://doi.org/10.1186/s13045-019-0712-y).

***Title: Gene Expression Indicates Altered Immune Modulation and Signaling Pathway Activation in Ovarian Cancer Patients Resistant to Topotecan**

Epithelial ovarian cancer (EOC) is one of the deadliest gynecological malignancies. Topotecan remains an essential tool in second-line therapy; even so, most patients develop resistance within a short period of time. We aimed to identify biomarkers of topotecan resistance by using gene expression signatures derived from patient specimens at surgery and available subsequent responses to therapy. Gene expression was collected for 1436 patients and 10,103 genes. Based on disease progression, patients were categorized as responders/nonresponders depending on their progression free survival (PFS) state at 9, 12, 15 and 18 months after surgery. For each gene, the median expression was compared between responders and nonresponders for two treatment regimens (chemotherapy including/excluding topotecan) with Mann–Whitney U test at each of the four different PFS cutoffs. Statistical significance was accepted in the case of $p < 0.05$ with a fold change (FC) ≥ 1.44. Four genes (EPB41L2, HLA-DQB1, LTF and SFRP1) were consistently overexpressed across multiple PFS cutoff times in initial tumor samples of patients with disease progression following topotecan treatment. A common theme linked to topotecan resistance was altered immune modulation. Genes associated with disease progression after systemic chemotherapy emphasize the role of the initial organization of the tumor microenvironment in therapy resistance. Our results uncover biomarkers with potential utility for patient stratification. The results of this study were published in International Journal of Molecular Sciences (https://doi.org/10.3390/ijms20112750).

***Title: Mutations Defining Patient Cohorts With Elevated PD-L1 Expression in Gastric Cancer**

The immunotherapy agent pembrolizumab has been approved for gastric cancer (GC) patients with recurrent or advanced disease who are PD-L1 positive. Mutations in the primary lesion may drive the expression of immune targets thereby priming the tumor to therapeutic sensitivity. In this study, we aimed to uncover mutations associated with elevated PD-L1 expression in GC patients. Data from 410 GC patients were available, including the mutational spectrum of 39,916 genes and expression values of 20,500 genes. PD-L1 gene

expression was compared to the mutational status of each gene separately by using a Mann-Whitney U-test and a Receiver Operating Characteristic test. Only mutations with a prevalence over 5% were considered. Significance was accepted in cases of p < 1E-05 and a fold change over 1.44. Mutations in 209 genes were associated with increased PD-L1 expression. These mutations were enriched in genes related to microtubule-based movement (p = 3.4E-4), cell adhesion (p = 4.9E-4), response to DNA-damage (p = 6.9E-4), and double-strand break-repair (p = 1.6E-3). Mutations in TTK (p = 8.8E-10, AUC = 0.77), COL7A1 (p = 2.0E-9, AUC = 0.74), KIF15 (p = 2.5E-9, AUC = 0.75), and BDP1 (p = 3.3E-9, AUC = 0.74) had the strongest link to elevated PD-L1 expression. Finally, we established a decision tree based on mutations in PIK3CA, MEF2C, SLC11A1, and KIF15 capable to separate patient sub-cohorts with elevated PD-L1 expression. In summary, we identified mutations associated with elevated PD-L1 expression that facilitate the development of better prognostic biomarkers for GC, and might offer insight into the underlying tumor biology. The results of this study were published in Frontiers in Pharmacology (https://doi.org/10.3389/fphar.2018.01522).

***Title: Uncovering Potential Therapeutic Targets in Colorectal Cancer by Deciphering Mutational Status and Expression of Druggable Oncogenes**

Background: Numerous driver mutations have been identified in colorectal cancer (CRC), but their relevance to the development of targeted therapies remains elusive. The secondary effects of pathogenic driver mutations on downstream signaling pathways offer a potential approach for the identification of therapeutic targets. We aimed to identify differentially expressed genes as potential drug targets linked to driver mutations. Methods: Somatic mutations and the gene expression data of 582 CRC patients were utilized, incorporating the mutational status of 39,916 and the expression levels of 20,500 genes. To uncover candidate targets, the expression levels of various genes in wild-type and mutant cases for the most frequent disruptive mutations were compared with a Mann–Whitney test. A survival analysis was performed in 2100 patients with transcriptomic gene expression data. Up-regulated genes associated with worse survival were filtered for potentially actionable targets. The most significant hits were validated in an independent set of 171 CRC patients. Results: Altogether, 426 disruptive mutation-associated upregulated genes were identified. Among these, 95 were linked to worse recurrence-free survival (RFS). Based on the druggability filter, 37 potentially actionable targets were revealed. We selected seven genes and validated their expression in 171 patient specimens. The best independently validated combinations were DUSP4 (p = 2.6 × 10−12) in ACVR2A mutated (7.7%) patients; BMP4 (p

= 1.6 × 10−04) in SOX9 mutated (8.1%) patients; TRIB2 (p = 1.35 × 10−14) in ACVR2A mutated patients; VSIG4 (p = 2.6 × 10−05) in ANK3 mutated (7.6%) patients, and DUSP4 (p = 7.1 × 10−04) in AMER1 mutated (8.2%) patients. Conclusions: The results uncovered potentially druggable genes in colorectal cancer. The identified mutations could enable future patient stratification for targeted therapy. The results of this study were published in Cancers (https://doi.org/10.3390/cancers11070983).

Title: JAK–STAT inhibition impairs K-RAS-driven lung adenocarcinoma progression

Oncogenic K-RAS has been difficult to target and currently there is no K-RAS-based targeted therapy available for patients suffering from K-RAS-driven lung adenocarcinoma (AC). Alternatively, targeting K-RAS-downstream effectors, K-RAS-cooperating signaling pathways or cancer hallmarks, such as tumor-promoting inflammation, has been shown to be a promising therapeutic strategy. Since the JAK–STAT pathway is considered to be a central player in inflammation-mediated tumorigenesis, we investigated here the implication of JAK–STAT signaling and the therapeutic potential of JAK1/2 inhibition in K-RAS-driven lung AC. Our data showed that JAK1 and JAK2 are activated in human lung AC and that increased activation of JAK–STAT signaling correlated with disease progression and K-RAS activity in human lung AC. Accordingly, administration of the JAK1/2 selective tyrosine kinase inhibitor ruxolitinib reduced proliferation of tumor cells and effectively reduced tumor progression in immunodeficient and immunocompetent mouse models of K-RAS-driven lung AC. Notably, JAK1/2 inhibition led to the establishment of an antitumorigenic tumor microenvironment, characterized by decreased levels of tumor-promoting chemokines and cytokines and reduced numbers of infiltrating myeloid derived suppressor cells, thereby impairing tumor growth. Taken together, we identified JAK1/2 inhibition as promising therapy for K-RAS-driven lung AC. The results of this study were published in International Journal of Cancer (https://doi.org/10.1002/ijc.32624).

***Title: Currently favored sampling practices for tumor sequencing can produce optimal results in the clinical setting**

Tumor heterogeneity is a consequence of clonal evolution, resulting in a fractal-like architecture with spatially separated main clones, sub-clones and single-cells. As sequencing an entire tumor is not feasible, we ask the question whether there is an optimal clinical sampling strategy that can handle heterogeneity and hypermutations? Here, we tested the effect of sample size, pooling strategy as well as sequencing depth using whole-exome

sequencing of ovarian tumor specimens paired with normal blood samples. Our study has an emphasis on clinical application—hence we compared single biopsy, combined local biopsies and combined multi-regional biopsies. Our results show that sequencing from spatially neighboring regions show similar genetic compositions, with few private mutations. Pooling samples from multiple distinct regions of the primary tumor did not increase the overall number of identified mutations but may increase the robustness of detecting clonal mutations. Hypermutating tumors are a special case, since increasing sample size can easily dilute sub-clonal private mutations below detection thresholds. In summary, we compared the effects of sampling strategies (single biopsy, multiple local samples, pooled global sample) on mutation detection by next generation sequencing. In view of the limitations of present tools and technologies, only one sequencing run per sample combined with high coverage (100–300 ×) sequencing is affordable and practical, regardless of the number of samples taken from the same patient. The results of this study were published in Scientific reports (https://doi.org/10.1038/s41598-020-71382-3).

***Title: Research funding: past performance is a stronger predictor of future scientific output than reviewer scores**

This and the next project are not related to the oriviginal research plan. Scientific grants are awarded almost exclusively on the basis of an independent peer review of a proposal submitted by the principal investigator (PI). The writing and reviewing of these applications consumes a significant amount of researchers' time. Here, we perform a large-scale performance evaluation of review-based grant allocation via analysis of the grant proposals submitted to the Hungarian Scientific Research Fund. In total, 42,905 scored review reports prepared for 13,303 proposals submitted between 2006 and 2015 were analyzed. The publication and citation characteristics of the PIs were obtained from the Hungarian Scientific Work Archive (www.mtmt.hu). Each publication was assigned to its respective SCImago Journal Rank category, and only publications in the first quarter (Q1) were considered. Citation, H-index and publication data were derived for each analyzed year for each researcher. Of all proposals, 3455 were funded (26%). PIs with a funded proposal had significantly more Q1 articles and first/last authored Q1 articles (1.91 vs. 1.30, $p<1e-16$ and 0.82 vs 0.53, $p<1e-16$, respectively). Of the successful applications, those involving international collaborations and extended budget had higher publication output. Applicant age, grant duration, and submission year were not correlated with publication performance. Reviewer scores displayed a minor association (corr.coeff = 0.08-011) with the number of Q1 publications. International reviewers were significantly less efficient than national reviewers

(p = 0.021). A strong correlation with output was observed for the scientometric characteristics of the applying PI at the time of submission, including H-index (corr.coeff = 0.45-0.54), independent citation (corr.coeff. = 0.46-0.62), and yearly average Q1 articles (corr.coeff = 0.63-0.79, p<1e-16). Similar correlations were observed for nonfunded applicants. We performed a comprehensive evaluation of review-based resource allocation efficiency in basic research funding. Evidence suggests that the past scientometric performance of the principal investigator is the best predictor of future output. These results were published in Journal of Infometrics (https://doi.org/10.1016/j.joi.2020.101050).

***Title: Is there a golden age in publication activity?—an analysis of age-related scholarly performance across all scientific disciplines**

We examined whether the publication characteristics of various scientific disciplines exhibit age-related trends. Our analysis was based on two large data sets comprising all major scientific disciplines. Citation data for European Research Council grant holders (ERC, n = 756) were obtained from Google Scholar. Publication data for Hungarian researchers (HUN, n = 2469) were obtained from the Hungarian Scientific Work Archive. The evaluated performance parameters include the number of citations received and the number of high quality first/last author papers published in the last five years. We designated the time between maximum growth and the achieved maximal annual value of total citations as the Golden Age of a researcher. Regarding citation growth, the mean age at the highest growth was 41.75 and 41.53 years for ERC grantees and Hungarian researchers, respectively. Each discipline had different values, with mathematics (38.5 years, ERC) and biology (34.7 years, HUN) having the youngest mean age of highest citation growth and agriculture (45.2 years, ERC) and language sciences (49.9 years, HUN) having the oldest mean age. The maximal growth of publications occurred at 44.5 years, with physics starting first (40.5 years, HUN) and language sciences as last (51.4 years, HUN). Most academic careers require decades to reach their peak and the length of the period of maximum performance varies across disciplines. The most creative time period is rising and is currently in the second half of the forties. Identifying the Golden Age in diverse research careers may be of substantial help in the distribution of grants and tenure positions. The results of this study were published in Scientometrics (https://doi.org/10.1007/s11192-020-03501-w).

***Title: Molecular stratifications, biomarker candidates and new therapeutic options in current medulloblastoma treatment approaches**

Medulloblastoma (MB) is the most common malignant childhood tumor of the brain. Multimodal treatment consisting of surgery, radiation therapy, and chemotherapy reduced cumulative incidence of late mortality but increased the incidence of subsequent neoplasms

and severe, incapacitating chronic health conditions. Present treatment strategies fail to recognize heterogeneity within patients despite wide divergence in individual responses. The persistent mortality rates and serious side effects of non-targeted cytotoxic therapies indicate a need for more refined therapeutic approaches. Advanced genomic research has led to the accumulation of an enormous amount of genetic information and resulted in a consensus distinguishing four molecular subgroups, WNT-activated, SHH-activated, and Group 3 and 4 medulloblastomas. These have distinct origin, demographics, molecular alterations, and clinical outcomes. Although subgroup affiliation does not predict response to therapy, new subgroup-specific markers of prognosis can enable a more layered risk stratification with additional subtypes within each primary subgroup. Here, we summarize subgroup-specific genetic alterations and their utility in current treatment strategies. The transition toward molecularly targeted interventions for newly diagnosed MBs remains slow, and prospective trials are needed to confirm stratifications based on molecular alterations. At the same time, numerous studies focus at fine-tuning the intensity of invasive radio- and chemotherapies to reduce intervention-related long-term morbidity. There are an increasing number of immunotherapy-based treatment strategies including immune checkpoint-inhibitors, oncolytic viruses, CAR-T therapy, and NK cells in recurrent and refractory MBs. Although most trials are in early phase, there is hope for therapeutic breakthroughs for advanced MBs within the next decade. The results of this study were published in Cancer and Metastasis Reviews (https://doi.org/10.1007/s10555-020-09854-1).

Title: Genome-wide alterations of uracil distribution patterns in human DNA upon chemotherapeutic treatments

Numerous anti-cancer drugs perturb thymidylate biosynthesis and lead to genomic uracil incorporation contributing to their antiproliferative effect. Still, it is not yet characterized if uracil incorporations have any positional preference. Here, we aimed to uncover genome-wide alterations in uracil pattern upon drug treatments in human cancer cell line models derived from HCT116. We developed a straightforward U-DNA sequencing method (U-DNA-Seq) that was combined with in situ super-resolution imaging. Using a novel robust analysis pipeline, we found broad regions with elevated probability of uracil occurrence both in treated and non-treated cells. Correlation with chromatin markers and other genomic features shows that non-treated cells possess uracil in the late replicating constitutive heterochromatic regions, while drug treatment induced a shift of incorporated uracil towards segments that are normally more active/functional. Data were corroborated by colocalization studies via dSTORM microscopy. This approach can be applied to study the dynamic spatio-temporal nature of genomic uracil. This paper was published in eLife (https://doi.org/10.7554/eLife.60498).

Title: Genomic Mapping Identifies Mutations in RYR2 and AHNAK as Associated with Favorable Outcome in Basal-Like Breast Tumors Expressing PD1/PD-L1

Treatment with anti-PD-L1 antibodies has shown efficacy in basal-like breast cancer. In this context, identification of pre-activated immune tumors is a main goal. Here we explore mutations in PD1 and PD-L1 high-expressing tumors to identify genomic correlates associated with outcome. To do so, RNA-seq and mutation data from 971 breast cancer patients from the TCGA dataset were used to identify most prevalent mutations in patients with high levels of PD1 and PD-L1. Transcriptomic signatures associated with the selected mutations were identified and analyzed in terms of outcome and immune cell infiltration. We identified co-occurrent mutations in RYR2 and AHNAK in 8% and 5% of basal-like tumors respectively, which conferred good prognosis in patients with high expression of PD1 and PD-L1 genes. The transcriptomic signature associated with these mutations, composed of CXCL9, GBP5, C1QA, IL2RG, CSF2RB, IDO1 and LAG3 genes, also conferred good prognosis and correlated with immune infiltrations within the tumors. The joint signature classified patients with favorable relapse-free survival (HR: 0.28; CI: 0.2–0.38; p = 1.7 × 10−16) and overall survival (HR: 0.18; CI: 0.09–0.34; p = 6.8 × 10−9), showing a stronger prediction capacity than previous reported signatures. In conclusion, we describe two novel mutations and their transcriptomic signature, both associated with a favorable outcome and immune infiltrates in PD1 and PD-L1 high-expressing basal-like tumors.This project was published in Cancers (https://doi.org/10.3390/cancers12082243).

***Title: Predictive biomarkers of platinum and taxane resistance using the transcriptomic data of 1816 ovarian cancer patients**

Objective: The first-line chemotherapy for ovarian cancer is based on a combination of platinum and taxane. To date, no reliable predictive biomarker has been recognized that is capable of identifying patients with pre-existing resistance to these agents. Here, we have established an integrated database and identified the most significant biomarker candidates for chemotherapy resistance in serous ovarian cancer. Methods: Gene arrays were collected from the GEO and TCGA repositories. Treatment response was defined based on pathological response or duration of relapse-free survival. The responder and nonresponder cohorts were compared using the Mann-Whitney and receiver operating characteristic tests. An independent validation set was established to investigate the correlation between chemotherapy response for the top 8 genes. Statistical significance was set at $p < 0.05$. Results: The entire database included 1816 tumor samples from 12 independent datasets. From analyzing all the genes for platinum + taxane response, we identified the eight strongest genes correlated to chemotherapy resistance: AKIP1 (p = 1.60E-08, AUC = 0.728),

MARVELD1 (p = 2.70E-07, AUC = 0.712), AKIRIN2 (p = 2.60E-07, AUC = 0.704), CFL1 (p = 8.10E-08, AUC = 0.694), SERBP1 (p = 8.10E-07, AUC = 0.684), PDXK (p = 1.30E-04, AUC = 0.634), TFE3 (p = 7.90E-05, AUC = 0.631) and NCOR2 (p = 1.90E-03, AUC = 0.611). Of these, the independent validation confirmed TFE3 (p = 0.012, AUC = 0.718), NCOR2 (p = 0.048, AUC = 0.671), PDXK (p = 0.019, AUC = 0.702), AKIP1 (p = 0.002, AUC = 0.773), MARVELD1 (p = 0.044, AUC = 0.675) and AKIRIN2 (p = 0.042, AUC = 0.676). An online interface was set up to enable future validation and ranking of new biomarker candidates in an automated manner (www.rocplot.org/ovar). Conclusions: We compiled a large integrated database with available treatment and response information and used this to uncover new biomarkers of chemotherapy response in serous ovarian cancer. The results of this study were published in Gynecologic Oncology (https://doi.org/10.1016/j.ygyno.2020.01.006).

***Title: ROCplot.org: Validating predictive biomarkers of chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3,104 breast cancer patients**

Systemic therapy of breast cancer can include chemotherapy, hormonal therapy and targeted therapy. Prognostic biomarkers are able to predict survival and predictive biomarkers are able to predict therapy response. In this report, we describe the initial release of the first available online tool able to identify gene expression-based predictive biomarkers using transcriptomic data of a large set of breast cancer patients. Published gene expression data of 36 publicly available datasets were integrated with treatment data into a unified database. Response to therapy was determined using either author-reported pathological complete response data (n = 1,775) or relapse-free survival status at 5 years (n = 1,329). Treatment data includes chemotherapy (n = 2,108), endocrine therapy (n = 971) and anti-human epidermal growth factor receptor 2 (HER2) therapy (n = 267). The transcriptomic database includes 20,089 unique genes and 54,675 probe sets. Gene expression and therapy response are compared using receiver operating characteristics and Mann–Whitney tests. We demonstrate the utility of the pipeline by cross-validating 23 paclitaxel resistance-associated genes in different molecular subtypes of breast cancer. An additional set of established biomarkers including TP53 for chemotherapy in Luminal breast cancer (p = 1.01E-19, AUC = 0.769), HER2 for trastuzumab therapy (p = 8.4E-04, AUC = 0.629) and PGR for hormonal therapy (p = 8.6E-05, AUC = 0.7), are also endorsed. The tool is designed to validate and rank new predictive biomarker candidates in real time. By analyzing the selected genes in a large set of independent patients, one can select the most robust candidates and quickly eliminate those that are most likely to fail in a clinical setting. The analysis tool is accessible at www.rocplot.org. The original publication of this study appeared in Internatioinal Journal of Cancer (https://doi.org/10.1002/ijc.32369).