Univerzita Karlova v Praze
1. lékařská fakulta

Charles University in Prague
First Faculty of Medicine

Autoreferát disertační práce
Summary of the Ph.D. Thesis

UNIVERZITA KARLOVA
1. lékařská fakulta

Výzkum klíčových mechanizmů onkogeneze s použitím modelových buněčných systémů

Investigating critical mechanisms of oncogenesis using cell model systems

Mgr. Hana Hušková
2017

Doktorské studijní programy v biomedicíně

Univerzita Karlova a Akademie věd České republiky

Obor: Fyziologie a patofyziologie člověka

Předseda oborové rady: Prof. MUDr. Jaroslav Pokorný, DrSc.

Školicí pracoviště: 1. lékařská fakulta Univerzity Karlovy v Praze, ústav Biocev, Vestec, Česká republika, a Inernational Agency for Research on Cancer (IARC), Lyon, Francie

Školitelé: Dr. Ing. Jiří Zavadil, Prof. MUDr. Tomáš Stopka Ph.D.

Disertační práce bude nejméně pět pracovních dnů před konáním obhajoby zveřejněna k nahlížení veřejnosti v tištěné podobě na Oddělení pro vědeckou činnost a zahraniční styky Děkanátu 1. lékařské fakulty.

Table of Contents

ABSTRAKT (CZ)

Lidé jsou v průběhu života vystaveni různým faktorům způsobujícím poškození DNA, vedoucí ke změnám v buněčné fyziologii a potenciálně k expanzi imortalizovaného buněčného klonu a vzniku nádoru. Mutace v DNA jsou jak záznamem o působení mutagenních procesů, tak klíčem k biologii a patofyziologii nádorů. Masivně paralelní sekvenování umožňuje sekvenování všech kódujících sekvencí či dokonce celých genomů lidských nádorů. Z těchto dat je možné získat vzorce mutací typické pro jednotlivé mutagenní procesy, stejně jako poukázat na mutace a geny hrající roli při vzniku a vývoji nádoru. Řada popsaných vzorců mutací však nemá známou příčinu a řada známých karcinogenů nemá dosud přiřazen mutační vzorec. Stejně tak se předpokládá, že dosud není známa řada mutací a genů s vlivem na vznik nádoru. Tato disertační práce charakterizuje experimentální systém založený na imortalizaci myších embryonálních fibroblastů (MEF) za působení mutagenu, umožňující určení vzorců mutací daných mutagenů a určení mutovaných genů důležitých pro vznik nádoru. Kultivace buněk MEF vede k jejich senescenci, která může být překonána mutacemi ve funkčně důležitých genech, analogicky ke stádiím vzniku lidských nádorů. Sekvenování kódujících sekvencí 25 imortalizovaných buněčných linií, které vznikly za působení rozličných mutagenů, ukázalo, že tento systém dokáže rekapitulovat vzorce mutací nalezené v lidských nádorech. Tyto buněčné linie také vykazovaly mutace v řadě genů důležitých pro vznik rakoviny u člověka a genů účastnících se epigenetické regulace. Skórovací systém, vyvinutý v rámci této práce, určil jako možné geny podporující vznik nádorů geny známé (např. Tp53 a Hras), ale i geny, jejichž vliv na vznik nádorů u člověka dosud nebyl zkoumán, jako je Smarcd2, kódující podjednotku komplexu BAF regulujícího chromatin. Použití molekulárního inhibitoru ukázalo, že MEF buněčná linie s mutací Smarcd2 je závislá na aktivitě komplexu PRC2, což koresponduje s výsledky získanými z lidských buněčných linií s mutacemi dalších podjednotek komplexu BAF. Předložená disertační práce ukazuje, že imortalizované linie z MEF buněk mohou být využity jako účinné modely pro studium důležitých aspektů vzniku nádorů.

Klíčová slova: mutace, vzorce mutací, mutagen, onkogen, tumor supresor, Ras, BAF

ABSTRACT (EN)

Humans and cells in their bodies are exposed to various mutagens in their lifetime that cause DNA damage and mutations, which affect the biology and physiology of the target cell, and can lead to the expansion of an immortalized cell clone. Genome-wide massively parallel sequencing allows the identification of DNA mutations in the coding sequences (whole exome sequencing, WES), or even the entire genome of a tumour. Mutational signatures of individual mutagenic processes can be extracted from these data, as well as mutations in genes potentially important for cancer development ('cancer drivers', as opposed to 'passengers', which do not confer a comparative growth advantage to a cell clone). Many known mutational signatures do not yet have an attributed cause; and many known mutagens do not have an attributed signature. Similarly, it is estimated that many cancer driver genes remain to be identified. This Thesis proposes a system based on immortalization of mouse embryonic fibroblasts (MEF) upon mutagen treatment for modelling of mutational signatures and identification and testing of cancer driver genes and mutations. The signatures extracted from WES data of 25 immortalized MEF cell lines, which arose upon treatment with a variety of mutagens, showed that the assay recapitulates the signatures of these compounds found in human tumours. The cell lines also harboured numerous mutations in genes known to act as cancer drivers in certain contexts, as well as mutations in a list of genes implicated in regulation of the epigenome. A scoring system devised for this study identified multiple putative drivers of the cancer-like phenotype of the cell lines, both well-known drivers (Tp53, Hras) as well as yet unrecognized putative ones (Smarcc1, Smarcd2 subunits of the BAF chromatin remodeling complex). Experiments using a small molecule inhibitor showed that the Smarcd2 mutation is likely to create a dependency of the affected cells on the PRC2 complex, as was previously demonstrated for other mutations in the BAF complex subunits in human cancer cell lines. In summary, the data presented in this Thesis show that the MEF cell lines are an invaluable resource for studies of certain aspects of human cancer development.

Keywords: mutations, mutational signature, mutagen, cancer driver, Ras, BAF

1. INTRODUCTION

Cancer is a group of diseases which can originate from various cell types, have diverse risk factors as well as epidemiological and clinical characteristics. It is characterized by uncontrolled proliferation of cells which do not respect normal tissue organization and can invade distant sites in the body.

It was demonstrated that mutagenic compounds bind to the the deoxyribonucleic acid (DNA), the molecular substance of genetic information, and cause its damage and base alterations in the DNA, leading to pathological processes in the affected cell and, ultimately, expansion of an immortal cellular clone and formation of a tumour. Thus, mutations in the DNA are central to cancer development and physiology.

1.1. Alterations in DNA as a record of mutagenic processes

Mutations in the tumour DNA are a result of mutagenic processes operative during the tumour's lifetime. Most mutations are so called somatic mutations, originating and present in a specific somatic cell clone, as opposed to germline mutations which are present in all cells of the body since the conception. Mutations can arise due to environmental carcinogens, which alter the DNA directly, or indirectly, or by inherent errors of the DNA replication machinery. Various reports estimate that environmental risk factors are responsible for 60-90 % of cancer cases.[1,2]

Human tumours usually develop under the influence of multiple mutagenic processes, making it challenging to distinguish the individual contributions to cancer development. To address the problem, Alexandrov et al. took an advantage of the vast amounts of data on somatic mutations in cancer, generated by massively parallel sequencing of human tumours and available in public repositories. Using non-negative matrix factorization method, they decomposed the frequencies of single base substitutions in 96 classes (6 mutation types – C>A, C>G, C>T, T>A, T>C, T>G – in 16 possible trinucleotide sequence contexts) to specific patterns of mutations, termed 'mutational signatures'. The method was applied on somatic mutation data from more than 12,000 tumours from 40 cancer types and identified 30 patterns of mutations which were termed 'mutational signatures. Some signatures were attributed to mutagenic processes (both innate and environmental) based on the knowledge of epidemiology and mechanisms of action[3,4]. However, many signatures still do not have an attributed cause, and vice versa, many known carcinogens do not have an attributed signature. A systematic experimental approach is needed to provide the explicit link between a mutagenic process and a mutational signature. The most valuable will be systems based on genome-wide

massively parallel sequencing, since they provide enough detail with a relatively small investment, as opposed to single- and reporter-gene based approaches.[5-10]

1.2. Alterations in DNA as the causes and effectors of tumour physiology

Some mutations in the DNA can confer a selective growth advantage to a cell clone, which can lead to its expansion and result in the development of a tumour. Such mutations are called 'cancer driver mutations' and the genes affected by these mutations are called 'cancer driver genes'. Driver genes with gain-of-function mutations are called 'oncogenes' and driver genes with inactivating mutations are called 'tumour suppressor genes'. Over 600 genes were found to be implicated in cancer so far and are now included in the Cancer Gene Census, a manually-curated database of cancer driver genes (Futreal et al., 2004). Most tumours bear several driver mutations; however, the majority of mutations, both somatic and germline, do not have an effect on cell fitness. These are called passenger mutations. Discriminating the 'drivers' from the 'passengers' is one of the main interests of cancer research.

Many key drivers that are frequently mutated in various cancer types, such as the *RAS* genes, *BRAF* or *TP53*, were identified based on experimental approaches (cloning and cell transformation assays). However, genes with lower mutation frequencies can also shift normal cells towards the cancer phenotype. One challenge that remains is the identification of driver genes that are mutated with low frequency.[11,12]

Since the mutations in driver genes confer a selective growth advantage to a cell clone, a) the driver genes should be mutated more frequently than expected from the background mutation rate, and b) the mutations affecting the driver gene will have a functional impact. Computational approaches to identify cancer driver genes using the human tumour sequencing data are based on these notions; they perform well in identifying known, frequently-mutated drivers, but give very different predictions when it comes to genes mutated with lower frequency.[11,13,14] A method based not on the individual genes, but on a network approach performed better.[15] The driver genes produce proteins which function within protein complexes, pathways or networks. These higher-level units could be thus seen as the ultimate cancer driver events. Though the abovementioned studies provided a lot of information, it remains descriptive and lacking mechanistic insight.

1.3. Mouse embryonic fibroblasts as a model of cancer development

Mouse embryonic fibroblasts (MEF) are primary cells. They grow in culture until senescence, which is bypassed, and immortalized cell lines arise. The immortalization step is a bottle neck, resembling those involved in the development of human tumours. MEF immortalization is easier than that of human cells, because MEFs do not undergo replicative senescence (they express telomerase and possess long telomeres). Senescence in MEFs is driven by other types of DNA damage, like reactive oxygen species (ROS) due to the culture conditions, and it is overcome by mutations in functionally important genes. Here we propose the carcinogen-exposure MEF immortalization assay for both modelling of mutational signatures, and identifying and testing novel driver genes (Figure 1).

2. HYPOTHESES AND AIMS OF THE THESIS

2.1. Hypotheses

I hypothesized that the carcinogen exposure and immortalization MEF assay recapitulates features relevant to the activity of used mutagens and selects for mutations that contribute to the cancer-like phenotype of the immortalized cells. Mutations act in a combinatorial manner and each clonal cell line results from the selection of a specific combination of growth-promoting driver mutations and driver genes. These can involve alterations in known, frequently-mutated genes as well as yet uncharacterized events. I further hypothesized that the driver mutations are introduced early in the assay due to the carcinogen treatment and are likely to become components of the carcinogen-specific mutational signature. Driver mutations can thus be identified from the pool of non-synonymous exposure-specific mutations with predicted functional impact, and tested in downstream validation experiments.

2.2. Aims of the Thesis

a) To generate **mutational signatures** of carcinogens using MEF immortalization assay, in order to recapitulate signatures observed in human tumour sequencing data.

b) To identify acquired mutations acting as **potential drivers** during immortalization of MEF cells.

c) To **functionally test** the impact and roles of select candidate driver mutations, both individually and in combination

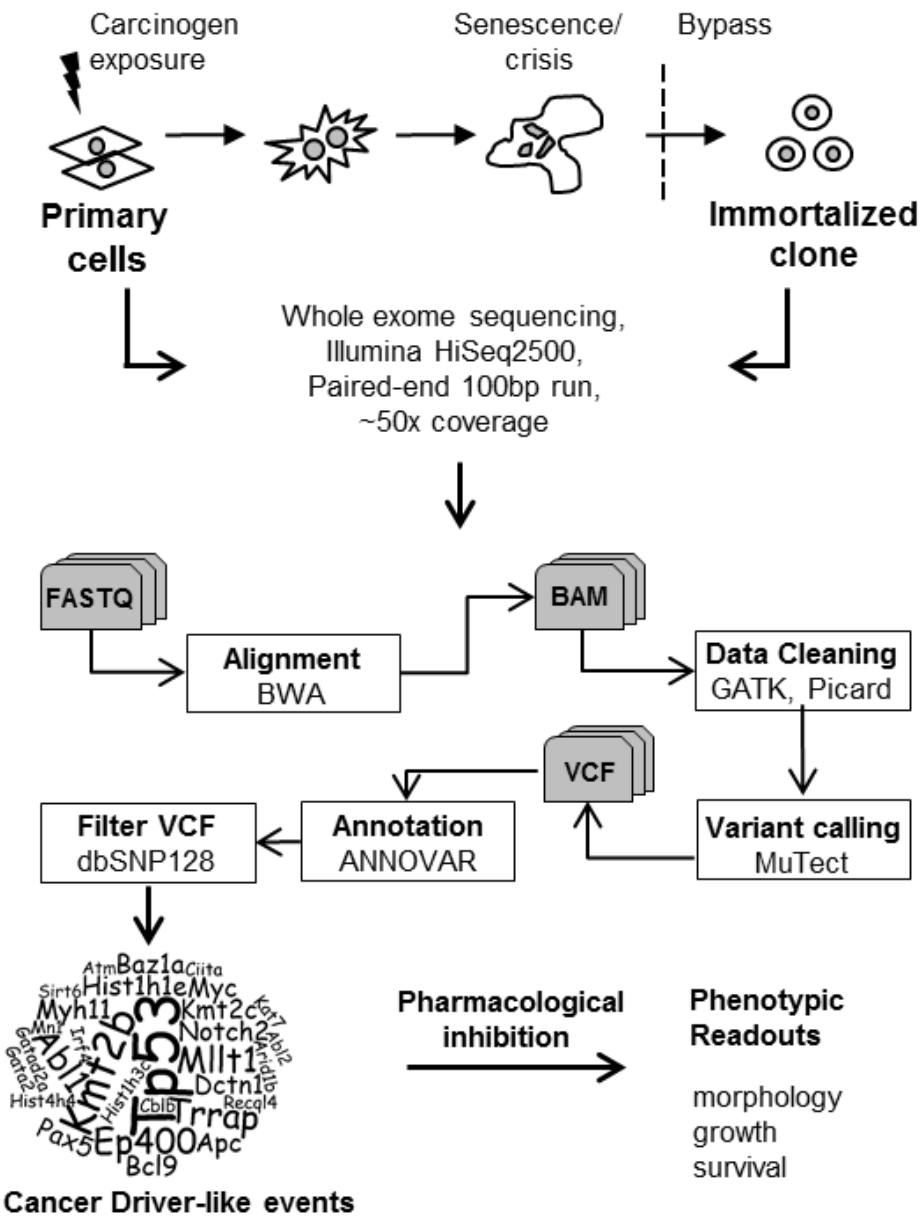


Figure 4: Study design. Mouse embryonic fibroblasts are exposed to a carcinogen in an early passage and cultivated until senescence and immortalization. Exome of the resulting cultures is sequenced at ~50× coverage. Data are analysed by the indicated pipeline and the results are used to extract mutational signatures and mine putative driver mutations. Impact of selected mutations is tested using small molecule inhibitors.

3. MATERIAL AND METHODS

3.1. Material

Twenty-six cell lines were used in the study (Table 1). Twenty-five were generated in the laboratory of Dr. Monica Hollstein in the German Cancer Research Center (DKFZ), Heidelberg, Germany, from the primary MEF cells of the Hupki (human p53 knock-in) mice which were exposed to various mutagens in an early passage, or Hupki crosses with a transgenic mouse expressing activation-induced cytidine deaminase (AID)[16]. The crosses were generated in the group of Prof. Hiroyuki Marusawa, Kyoto University, Japan. One cell line was generated by Hana Huskova at the International Agency for Research on Cancer (IARC), Lyon, France.

3.2. Methods

3.2.1. Whole exome sequencing and alignment

Library preparation and sequencing was outsourced to several companies, which applied comparable protocols, and sequenced the libraries using Illumina HiSeq2500 system. The reads were aligned to the mm9 mouse genome build using the BWA-MEM. Duplicate marking, realignment around indels and base recalibration was done using Picard and GATK, respectively. The mean depth-of-coverage was 54. Bam files were uploaded to the National Center for Biotechnology Information BioProjects web site, accession number PRJNA238303.

3.2.2. Variant calling, mutational signature analysis, pathway analysis

Variants were called with MuTect software using default parameters. Variants were annotated with ANNOVAR and single nucleotide polymorphisms according to the dbSNP database were filtered out. Mutational spectra and signature analysis was performed using the MutSpec toolbox in Galaxy.[17] Variants were filtered for exonic non-synonymous single base substitutions and splice site mutations. RefSeq-annotated genes affected by these variants were analysed using DAVID and IPA with relaxed criteria.

3.2.3. Driver gene identification and testing

Variants were filtered for exonic non-synonymous and splicing mutations and these were inspected for mutations in cancer-related genes and chromatin associated genes and regulators of the epigenome (Vogelstein et al., 2013, Gonzalez-Perez et al., 2013, Futreal et al., 2004). Mutations were prioritized using a simple scoring system based on their allelic frequency, mutation type, and predicted functional effect. Cell lines were subcloned and putative driver mutations were validated in individual clones by Sanger sequencing. Mutations in Ras genes

and BAF complex were chosen for testing. The effect of Ras mutations was tested using Mek inhibitor U0126, and the interplay between BAF and PRC2 complex was tested using Ezh2 inhibitor GSK126.

Table 1: Cell lines

Cell line ID	Exposure type	Exposure dose	Exposure duration	Origin
AA_1	AA	50 µM	4 days	DKFZ
AA_2	AA	50 µM	4 days	DKFZ
AA_3	AA	50 µM	4 days	DKFZ
AA_4	AA	50 µM	4 days	DKFZ
AA_5	AA	50 µM	4 days	DKFZ
AA_6	AA	50 µM	12 days	DKFZ
AA_7	AA	50 µM	8 days	DKFZ
AFB1_1	AFB1	2 µM	8 days	DKFZ
AFB1_2	AFB1	2 µM	8 days	DKFZ
AFB1_3	AFB1	2 µM	8 days	DKFZ
AID_1	None	n. a.	n. a.	DKFZ
AID_2	None	n. a.	n. a.	DKFZ
B[a]P_1	BaP	1 µM	6 days	DKFZ
B[a]P_2	BaP	1 µM	6 days	DKFZ
B[a]P_3	BaP	5 µM	2 days	DKFZ
MNNG_1	MNNG	20 µM	2 hours	DKFZ
MNNG_2	MNNG	20 µM	2 hours	DKFZ
MNNG_3	MNNG	20 µM	2 hours	DKFZ
MNNG_4	MNNG	20 µM	2 hours	DKFZ
Spont_1	None	n. a.	n. a.	DKFZ
Spont_2	None	n. a.	n. a.	DKFZ
Spont_3	None	n. a.	n. a.	DKFZ
Spont_4	None	n. a.	n. a.	DKFZ
Spont_5	None	n. a.	n. a.	IARC MMB
UVC_1	UVC	20 J/m^2	n. a.	DKFZ
UVC_2	UVC	20 J/m^2	n. a.	DKFZ

AA - aristolochic acid, AFB1 - aflatoxin B1, AID - activation-induced cytidine deaminase, B[a]P - benzo[a]pyrene, DKFZ - German Cancer Research Center, IARC MMB – Molecular Mechanisms and Biomarkers Group at the International Agnecy for Research on Cancer, MNNG - N-methyl-N'-nitro-N-nitrosoguanidine, n. a. – not applicable, Spont - spontaneous immortalization, UVC - ultraviolet light class C.

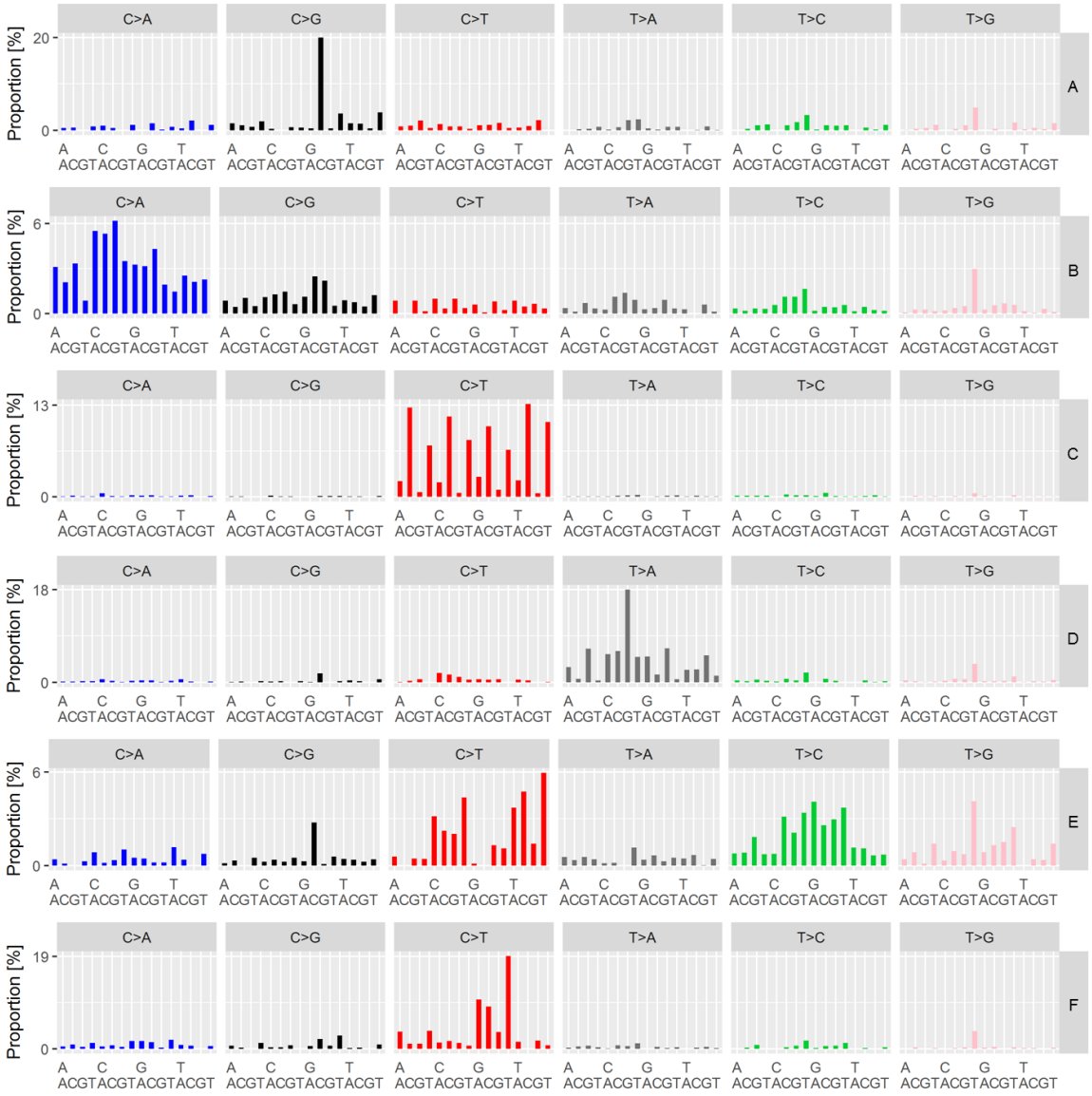
4. RESULTS

Twenty-six immortalized cell lines derived from primary Hupki mouse embryonic fibroblasts using modified 3T3 protocol were selected. Nineteen of the cell lines emerged after treatment with five different carcinogens: aristolochic acid (AA, N=7), aflatoxin B1 (AFB1, N=3), benzo[a]pyrene (B[a]P, N=3), N-methyl-N'-nitro-N-nitrosoguanidine (MNNG, N=4) and ultraviolet light class C (UVC, N=2). Two cell lines developed from primary Hupki MEFs which were engineered to overexpress AID, a DNA-mutating enzyme of the APOBEC family (N=2). Five cell lines immortalized spontaneously (Spont). Twenty-five cell lines were assigned to a test set for analyses, whereas one spontaneously immortalized cell line (Spont_5) was used as a control for subsequent experiments.

4.1.Mutational signatures analysis

Sequencing analysis of the 25 cell lines included in the test set yielded total of 16,061 single-base substitutions. The mutation load in the cell lines varied from ~100 to ~1,500 variants per cell line.

Six mutational signatures were extracted from the data using the NMF method and named A-F (Figure 2A). The main feature of the **signature A** was a pronounced peak of C>G mutations in 5'-G_C-3' context. This peak can be also spotted in signatures B, D, E and F. Signature A did not bear resemblance to any of the 30 COSMIC signatures, and it consisted mostly of mutations found in spontaneously immortalized cell lines, although many other cell lines contributed to it, too (Figure 2B). This signature is probably linked to culture conditions. **Signature B** displayed high frequency of C>A mutations in various sequence contexts, and consisted mostly of mutations found in B[a]P and AFB1 cell lines. Signatures which were mostly similar to signature B were signature 4 (tobacco smoking, similarity 0.82), signature 24 (aflatoxin, similarity 0.76) and signature 29 (tobacco chewing, similarity 0.72). Separate analyses showed that both AFB- and B[a]P-treated cell lines bear higher similarity to smoking signature, rather than to the AFB1 signature. **Signature C** consisted of C>T mutations in 5'-N_R-3' context (N – any base, R – pyrimidine). It was specific to cell lines derived from cells treated with alkylating agent MNNG. Signature C is identical to the COSMIC signature 11 (similarity 0.98), which has been attributed to exposure to the alkylating drug temozolomide. Similarly, **signature D** was identical to signature 22 (similarity 0.96), which has been linked to the exposure to aristolochic acids. The signature was rich in T>A mutations with a peak



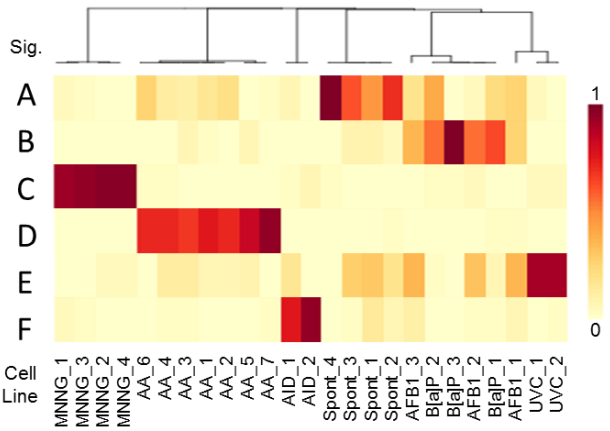


Figure 2: Mutational signatures identified in 25 immortalized MEF cell lines. A - Six mutational signatures were extracted using non-negative matrix factorization algorithm, based on frequency of 6 mutation types in 16 different sequence contexts. The signatures were named A-F. B - Relative contribution of mutations from individual MEF cell lines to mutational signatures A-F, identified in the pooled data set.

in 5‘-C_G-3‘ context. Signature D was specific to cell lines derived from cells exposed to AA. **Signature E** displayed high proportion of C>T and T>C mutations, and, to a lesser extent, T>G mutations. This signature was mostly composed of mutations detected in UVC-exposed cell lines, but many other cell lines also contributed to this signature. Signature E does not show a considerable similarity to any of the COSMIC signatures. **Signature F** was defined by a high proportion of C>T mutations with a noticeable peak in the 5‘-G_T-3’ context. This signature was specific to cell lines developed from cells overexpressing the AID transgene. It is not similar to any of the 30 COSMIC signatures. However, C>T mutations in 5‘-G_T-3‘, 5‘-G_A-3’ and 5‘-G_C-3’ contexts are typical for AID activity in the immunoglobulin gene (Rogozin and Kolchanov, 1992, Puente et al., 2015). Analogous signature was identified in whole genome sequencing data from 30 samples of chronic lymphocytic leukaemia and attributed to ectopic activity of AID (Kasar et al., 2015).

Together, these data indicate that MEF cell lines can recapitulate mutational signatures found in human cancers.

4.2. Functional annotation of mutations in the immortalized MEF cell lines

Pathway analysis of nonsynonymous and splicing mutations found in MEF cell lines was performed, to evaluate the functional effects of the mutagenic processes. Among the frequently affected pathways were those involved in structural integrity, regulation of cell cycle, proliferation, apoptosis and differentiation, as well as chromatin modification and transcription regulation. The processes, which were found frequently affected in the set of 25 immortalized MEF cell lines, are also frequently deregulated in human tumours and can be classified under the hallmarks of cancer[18]. These results indicate that processes affected by mutations in human cancers are also affected in immortalized MEF cell lines.

We next focused on mutations in cancer genes (as listed in the Cancer Gene Census) and genes involved in regulation of the epigenome, which have been recently identified to be frequently mutated in cancer.[15,19,20] More than 300 hits were found by filtering nonsynonymous exonic and splicing mutations against the Cancer Gene Census, many of the genes were mutated recurrently (Figure 3), marking a potential selection for mutations in these genes. Furthermore, we found 105 nonsynonymous and splicing mutations in 66 epigenome regulators, as well as 34 mutations in 23 histone genes. Twenty epigenetic modifiers and four histone genes were mutated recurrently (Figure 3).

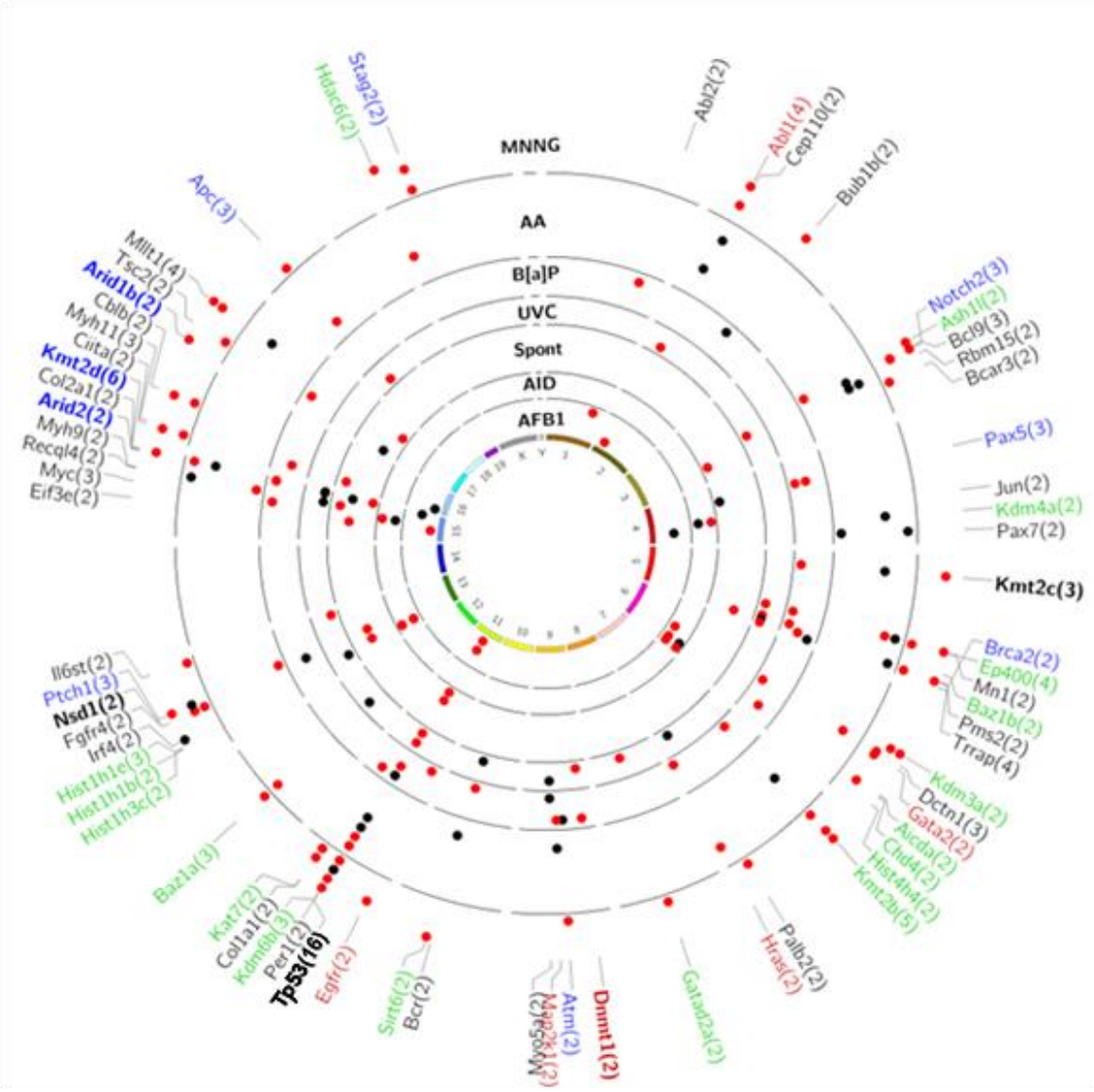


Figure 3: Recurrently mutated cancer and epigenetic modifier genes in 25 Hupki MEF cell lines. Genes listed in the Cancer Gene Census (black)[21], oncogenes (red) and tumor suppressor genes[22] (blue) and epigenetic modifiers[19] and histone genes (green) are indicated. Epigenetic modifiers that are also listed in the Cancer Gene Census are indicated in bold black. Epigenetic modifiers that are also listed as tumor suppressor genes[22] are in bold blue. Epigenetic modifiers that are also listed as oncogenes[22] are in bold red. Cell lines are arranged concentrically and grouped by carcinogen exposure. Red and black dots represent exposure-predominant and exposure non-predominant mutation types, respectively.

4.3. Patterns of mutations in protein complexes regulating the epigenome

The set of cell lines contained 9 mutations in the subunits of the ATP-dependent chromatin remodelling BAF complex. The mutations displayed a mutually-exclusive pattern, as was published previously, and validated by our analysis of human tumour sequencing data. Another almost mutually exclusive pattern observed in the MEF cell lines were mutations in Ep400 and Trrap subunits of the histone acetylase TIP60 complex (Figure A), and was validated

in the human tumour sequencing data (Figure 4B). This novel finding underpins the utility of MEFs for studying the human tumour biology.

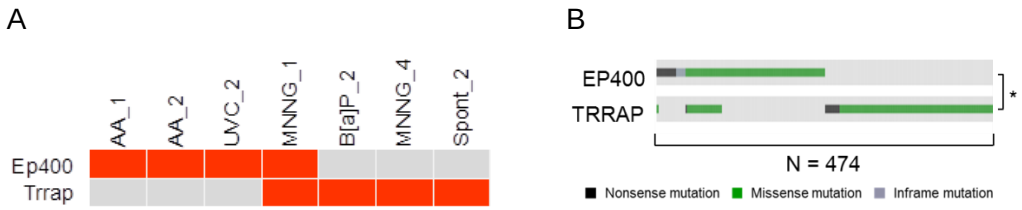


Figure 4: Analysis of Trrap and Ep400 mutations in mouse and human samples. A - TIP60 complex subunits Ep400 and Trrap mutated in MEF BBCE cell lines. B - TIP60 complex subunits EP400 and TRRAP mutated in human sequencing studies included in cBioPortal[23,24]. Result of X^2-test indicated, *p<0.001.

4.4. Functional testing of selected cancer driver genes/mutations

Putative driver mutations were selected based on the scoring system devised for this purpose. Among the high-scoring mutations were wel-known driver mutations ($Hras^{Q61L}$, $Tp53^{N131Y}$)[25,26], as well as mutations in genes not yet linked to cancer in human tumour sequencing studies, for example the BAF complex subunits ($Smarcc1^{H119L}$, $Smarcd2^{G166E}$). After validating the effect of activating Hras mutation in the MEF cell lines, we set to examine its interplay with BAF complex mutations.

A previous elegant study has demonstrated that cancer cell lines with mutations in the BAF chromatin remodelling complex cause a dependency of the cells to the function of the PRC2 histone methyltransferase complex[27]. Inhibiting the function of the PRC2 complex in BAF-mutant cell lines lead to cell death. However, the effect was attenuated in cell lines with a concomitant Ras mutation. The same was observed in the MEF cell lines (Figure 5). AA_2-1 cell clone ($Hras^{Q61L}$, $Smarcc2^{H119L}$) was the most resistant to the treatment with the inhibitor of the PRC2 catalytic subunit EZH2, as shown by both MTS and colony formation assays (Figure 5A,B). In contrary, the MNNG_4-2 clone ($Smarcd2^{G166E}$) was highly sensitive to the inhibitor and did not show any remaining viability after 3 days (MTS assay) and 7 days (colony formation assay), respectively. The cell line Spont_5, which was wild type for both Ras and BAF, was more sensitive to the inhibitor than the AA_2-1 clone, but a fraction of cells survived until the end of both MTS and colony formation assay experiments. Importantly, on the molecular level, the inhibitor was comparably effective in all tested cultures, as measure by its target H3K27me3 (Figure 5C), and the effects must therefore be attributed to the variable sensitivity of the cultures to PRC2 inhibition. The experiment was biologically validated

in another set of MEF cell clones with BAF mutations, either alone (MNNG_1-1, AFB1_3-2, BaP_1-2), or in combination with an activating Ras mutation (UVC_2-3). In summary, these results show that MEF immortalization assay selects for, and allows the testing of mutations important for human tumour physiology.

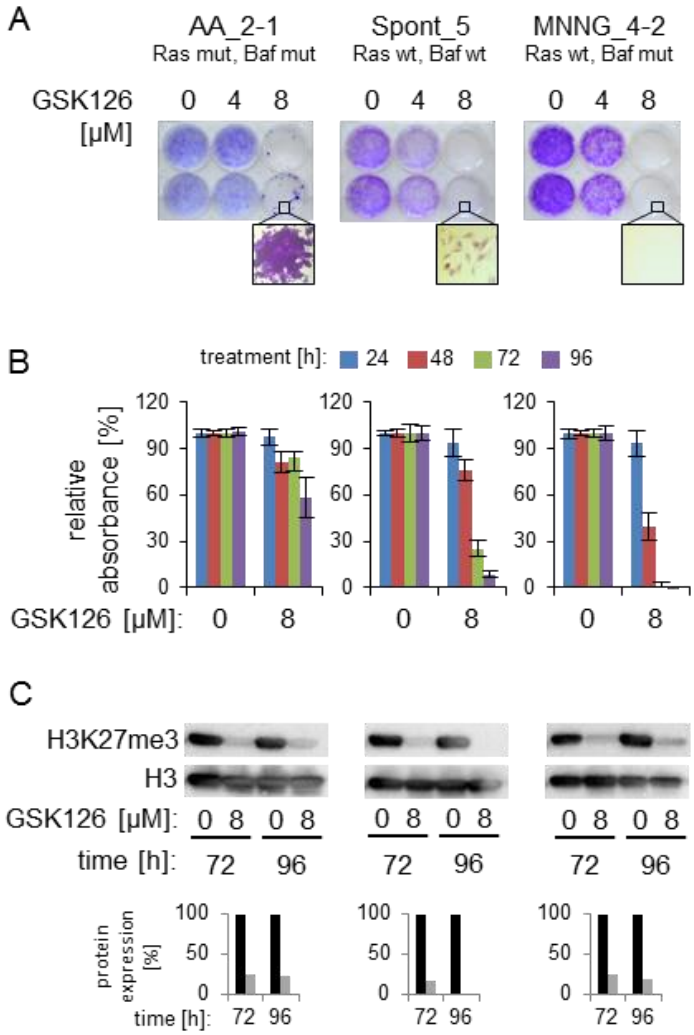


Figure 5: Effect of Ezh2 inhibitor treatment – first set of experiments. A – results of colony formation assay. Cells were seeded in a low density and treated with Ezh2 inhibitor GSK126, or with carrier (DMSO). Colonies were visualized after 7 days using crystal violet staining. Window shows 100x magnification. B – results of MTT assay. Cells were treated with Ezh2 inhibitor or carrier (DMSO) and absorbance was measured at indicated time points. Results of three independent experiments are plotted as mean and standard error of mean. C – Immunoblot for H3K27me3 mark in cells treated with Ezh2 inhibitor and a carrier. H3 was used as loading control. Abundance of H3K27me3 is plotted (treated relative to untreated cells).

5. DISCUSSION

The results of this proof-of-principle study show that the MEF immortalization assay coupled with genome-wide sequencing generates enough data for the extraction of mutational signatures using the NMF method. It recapitulates mutation signatures of environmental mutagenic compounds (AA, MNNG, B[a]P) and innate mutagenic processes (AID) found in human cancer[4,28,29]. The signature extracted from the AFB1-treated cell lines was more similar to the smoking signature than to the aflatoxin signature from the COSMIC database. This could be due to differences between mouse and human metabolism or DNA repair, and/or more complex aflatoxin composition in the real-life human exposure (not only AFB1, but also other types of aflatoxins). MEF UVC signature did not resemble any COSMIC signature. Potentially, the UVC functions differently than the common UVA and UVB which are not, like UVC, absorbed by the atmosphere and, unlike UVC, contribute to development of human cancer.

Hupki MEF immortalization coupled with massively parallel sequencing was one of the first approaches allowing modelling mutational signatures of human cancers using mammalian cells[30]. Other systems like human renal tubule HK-2 cell line[31], human mammary epithelial cells HMEC[32,33] provide useful data from human systems. The real asset of these systems is that they produce mutational signatures specific for the cells which are the targets of the compounds used in the assays (AA in the case of HK-2, B[a]P in the case of HMEC). However, MEFs produce correct mutational signatures – at least for the compounds described in the Thesis –, are much easier to handle compared to the abovementioned human cells, and the assay is relatively short (2 months vs. 6 months for human-cell systems). The current state of research on mutational signatures, and experimental systems which utilized to study them, was recently summarized by Hollstein et al.[34] and Zhivagui et al.[35]

Furthermore, the results described above show that MEF immortalization assay selects for mutations in genes important for cancer development and directly allows testing them on the corresponding mutational background. Though human tumour sequencing studies usually include an analysis of driver genes, they either entirely lack experimental validation[28,36,37], or it is done in a tumour cell line, on a different mutational background than on which the mutation operated in the tumour where it was discovered.[29,38] This is important, since the various dependencies may change the outcome completely. An important outcome of the Thesis is

a system for evaluation of putative driver mutations in MEF cell lines which developed upon treatment with a mutagen.

Surely, there are limitations in terms of the type of research to which the MEF cell lines can serve. Tumours are kind of organs of themselves. They start from a cell clone with specific properties, but as they develop, they have to organize and adapt (e.g. induce vascularization to get enough nutrition, manipulate immune cells to evade immune response). In the end, a tumour is composed not only from the tumorigenic cell clone, but also from other cell types which contribute to the tumour development. MEF cell line assay in its current state cannot be used to study the interplay between different cell types in the tumour. However, it can provide an insight into the tumour-founding clone biology: cell cycle and apoptosis alterations, epigenetics, metabolism, migration properties.

6. CONCLUSIONS

a. Whole exome sequencing of immortalized MEF cell lines produced sufficient data to extract mutational signatures. MEF mutational signatures of AA and MNNG closely resembled their corresponding signatures extracted from human tumours, as did the MEF mutational signatures of AID and B[a]P. The MEF mutational signature of AFB1 did not closely resemble the one extracted from liver tumours, which could be due to differences between mouse and human metabolism or DNA repair and/or more complex aflatoxin composition in the real-life human exposure. MEF UVC and spontaneous signatures did not resemble any COSMIC signature.

b. MEF cell lines bore nonsynonymous mutations in numerous genes included in the Cancer Gene Census, as well as in many genes involved in regulation of the epigenome. Furthermore, MEF cell lines displayed cancer-like mutation profile in terms of affected pathways, as well as alterations in the BAF and TIP60 chromatin-modifying complexes. Subunits of the BAF complex were mutated in 9 out of 25 MEF cell lines in a mutual exclusive manner, as earlier described in human cancers.[15,19] Similarly, Ep400 and Trrap subunits of the TIP60 acetyltransferase complex were nearly mutually exclusively mutated in 7 out of 25 MEF cell lines. The Ep400/Trrap pattern was confirmed in sequencing data from human tumours, describing for the first time this feature of human cancers.

c. Driver mutations can be identified in MEF cell lines. I devised a scoring system to identify mutations potentially driving the cancer-like phenotype of MEF cell lines among the wealth of mutations identified by WES. The scoring system is based on mutation type, allelic frequency and prediction of functional effect of the mutation. When applied on the Cancer Gene Census and epimodifier genes from two cell lines, the algorithm identified known Tp53 and Ras driver mutations, but also mutations in other BAF complex subunits, which were not previously identified as driver genes in human tumour sequencing studies.

d. MEF cell lines, in contrary to human tumours, permit in vitro manipulations, and thus functional testing of putative driver mutations. Inhibition of Ras signalling by Mek inhibitor in Ras-mutant cell lines lead to decreased cell viability in serum-deprived medium, while this effect was not observed for Ras wild type cell lines. Inhibition of Ezh2 activity in BAF-mutant cell lines lead to elimination of the cells, in contrary to Ras-BAF double mutants. This was demonstrated earlier in human cancer cell lines[27]. We extend the list of BAF complex mutations conferring a vulnerability to Ezh2 inhibition for Smarcd2^{G166E}.

7. REFERENCES

1. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science.* 2017;355(6331):1330-1334.
2. Wu S, Powers S, Zhu W, Hannun YA. Substantial contribution of extrinsic risk factors to cancer development. *Nature.* 2015.
3. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* 2013;3(1):246-259.
4. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-421.
5. Besaratinia A, Pfeifer GP. Applications of the human p53 knock-in (Hupki) mouse model for human carcinogen testing. *FASEB J.* 2010;24(8):2612-2619.
6. Besaratinia A, Li H, Yoon JI, Zheng A, Gao H, Tommasi S. A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens. *Nucleic Acids Res.* 2012;40(15):e116.
7. Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. *Mutat Res.* 2005;571(1-2):19-31.
8. Levy DD, Groopman JD, Lim SE, Seidman MM, Kraemer KH. Sequence specificity of aflatoxin B1-induced mutations in a plasmid replicated in xeroderma pigmentosum and DNA repair proficient human cells. *Cancer Res.* 1992;52(20):5668-5673.
9. Sage E, Drobetsky EA, Moustacchi E. 8-Methoxypsoralen induced mutations are highly targeted at crosslinkable sites of photoaddition on the non-transcribed strand of a mammalian chromosomal gene. *EMBO J.* 1993;12(2):397-402.

10. Manjanatha MG, Guo LW, Shelton SD, Doerge DR. Acrylamide-induced carcinogenicity in mouse lung involves mutagenicity: cII gene mutations in the lung of big blue mice exposed to acrylamide and glycidamide for up to 4 weeks. *Environ Mol Mutagen.* 2015;56(5):446-456.
11. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013;499(7457):214-218.
12. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014;505(7484):495-501.
13. Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. Evaluating the evaluation of cancer driver genes. *Proc Natl Acad Sci U S A.* 2016;113(50):14330-14335.
14. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics.* 2013;29(18):2238-2244.
15. Leiserson MD, Vandin F, Wu HT, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet.* 2015;47(2):106-114.
16. Okazaki IM, Hiai H, Kakazu N, et al. Constitutive expression of AID leads to tumorigenesis. *J Exp Med.* 2003;197(9):1173-1181.
17. Ardin M, Cahais V, Castells X, et al. MutSpec: a Galaxy toolbox for streamlined analyses of somatic mutation spectra in human and mouse cancer genomes. *BMC Bioinformatics.* 2016;17:170.
18. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-674.
19. Gonzalez-Perez A, Jene-Sanz A, Lopez-Bigas N. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* 2013;14(9):r106.
20. Feinberg AP, Koldobskiy MA, Göndör A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet.* 2016;17(5):284-299.
21. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Rev Cancer.* 2004;4(3):177-183.
22. Vogelstein B, Papadopoulos N, Velculescu V, Zhou S, Diaz L, Kinzler K. Cancer Genome Landscapes. *Science.* 2013;339(6127):1546-1558.
23. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-404.
24. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269):pl1.
25. Shaw RJ, Cantley LC. Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature.* 2006;441(7092):424-430.
26. Odell AF, Odell LR, Askham JM, Alogheli H, Ponnambalam S, Hollstein M. A novel p53 mutant found in iatrogenic urothelial cancers is dysfunctional and can be rescued by a second-site global suppressor mutation. *J Biol Chem.* 2013;288(23):16704-16714.
27. Kim KH, Kim W, Howard TP, et al. SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat Med.* 2015;21(12):1491-1496.
28. Kasar S, Kim J, Improgo R, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun.* 2015;6:8866.
29. Puente XS, Beà S, Valdés-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature.* 2015;526(7574):519-524.
30. Olivier M, Weninger A, Ardin M, et al. Modelling mutational landscapes of human cancers in vitro. *Sci Rep.* 2014;4:4482.

31. Poon SL, Pang ST, McPherson JR, et al. Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci Transl Med.* 2013;5(197):197ra101.
32. Severson PL, Vrba L, Stampfer MR, Futscher BW. Exome-wide mutation profile in benzo[a]pyrene-derived post-stasis and immortal human mammary epithelial cells. *Mutat Res Genet Toxicol Environ Mutagen.* 2014;775-776:48-54.
33. Stampfer MR, Bartley JC. Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. *Proc Natl Acad Sci U S A.* 1985;82(8):2394-2398.
34. Hollstein M, Alexandrov LB, Wild CP, Ardin M, Zavadil J. Base changes in tumour DNA have the power to reveal the causes and evolution of cancer. *Oncogene.* 2017;36(2):158-167.
35. Zhivagui M, Korenjak M, Zavadil J. Modelling Mutation Spectra of Human Carcinogens Using Experimental Systems. *Basic Clin Pharmacol Toxicol.* 2016.
36. Hayward NK, Wilmott JS, Waddell N, et al. Whole-genome landscapes of major melanoma subtypes. *Nature.* 2017;545(7653):175-180.
37. Pereira B, Chin SF, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016;7:11479.
38. Guichard C, Amaddeo G, Imbeaud S, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet.* 2012;44(6):694-698.

8. LIST OF PUBLICATIONS

8.1 Related to the Thesis

Huskova H, Ardin M, Weninger A, Vargova K, Barrin S, Villar S, Olivier M, Stopka T, Herceg Z, Hollstein M, Zavadil J, Korenjak M. 2017. Modeling cancer driver events in vitro using barrier bypass-clonal expansion assays and massively parallel sequencing. Oncogene (accepted for publication). (**IF=7.932**)

Olivier M, Weninger A, Ardin M, **Huskova H**, Castells X, Vallée MP, McKay J, Nedelko T, Muehlbauer KR, Marusawa H, Alexander J, Hazelwood L, Byrnes G, Hollstein M and Zavadil J. 2014. Modelling mutational landscapes of human cancers *in vitro*. Sci Rep 4: 4482. (**IF=5.578**)

8.2 Unrelated to the Thesis

Huskova H, Korecka K, Karban J, Vargova J, Vargova K, Dusilkova N, Trneny M, Stopka T. 2015. Oncogenic microRNA-155 and its target PU.1: an integrative gene expression study in six of the most prevalent lymphomas. Int J Hematol 102: 441-450. (**IF=1.846**)