UNIVERSITY OF TURKU

NOVEL PP2A BIOMARKERS IN CANCER

Eleonora Mäkelä

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. D OSA - TOM. 1538 | MEDICA - ODONTOLOGICA | TURKU 2021

UNIVERSITY OF TURKU

NOVEL PP2A BIOMARKERS IN CANCER

Eleonora Mäkelä

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1538 | MEDICA – ODONTOLOGICA | TURKU 2021

University of Turku

Faculty of Medicine
Institute of Biomedicine
Pathology
Turku Doctoral Programme of Molecular Medicine
Turku Bioscience Centre, University of Turku and Åbo Akademi University

Supervised by

Professor Jukka Westermarck, M.D., Ph.D.
Institute of Biomedicine, University of Turku
Turku Bioscience Centre, University of Turku
and Åbo Akademi University
Turku, Finland

Reviewed by

Professor María D. Odero, M.D., Ph.D.
Department of Biochemistry and Genetics
University of Navarra
Navarra, Spain

Caroline Heckman, Ph.D.
Institute for Molecular Medicine
Finland – FIMM
University of Helsinki
Helsinki, Finland

Opponent

Professor Kimmo Porkka, M.D., Ph.D.
Helsinki University Hospital Comprehensive
Cancer Center and Hematology Research
Unit Helsinki
University of Helsinki
Helsinki, Finland

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-8392-6 (PRINT)
ISBN 978-951-29-8393-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama Oy, Turku, Finland 2021

To my family

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Pathology
Turku Bioscience Centre
ELEONORA MÄKELÄ: Novel PP2A biomarkers in cancer
Doctoral Dissertation, 210 pp.
Turku Doctoral Programme of Molecular Medicine
March 2021

ABSTRACT

Inhibition of tumor suppressor Protein Phosphatase 2A (PP2A) has been denoted as one of the minimal requirements for malignant transformation of a human cell. In cancer, PP2A activity is in the majority of cases inhibited by the overexpression of the endogenous PP2A inhibitor proteins (PAIPs), such as CIP2A, ARPP19, SET, PME-1 and TIPRL. In this thesis, PP2A inhibition by the PAIPs was elaborated in acute and chronic myeloid leukemias (AML and CML) and head and neck squamous cell carcinoma (HNSCC). Moreover, the ability of the PAIPs to act as clinically relevant cancer biomarkers was investigated.

In HNSCC, radiotherapy is a mainstay for treatment, and thus it would be clinically very relevant to identify markers that would predict the radiation resistance of HNSCC tumors. In this thesis, we identified CIP2A as a novel OCT4 target gene. CIP2A and OCT4 were also found to be involved in HNSCC radioresistance. Our data propose that analysis of HNSCC tumors for OCT4 or OCT4/CIP2A double positivity at HNSCC diagnosis could be used to predict the radiation resistance of HNSCC tumors. These same targets could be further utilized in radiosensitation.

Despite a great number of genetic studies conducted on myeloid leukemias, the mechanisms that promote leukemia disease relapse and progression are not fully understood. In this thesis, ARPP19 was identified as a novel predictive relapse biomarker in AML. In AML cells, ARPP19 depletion resulted in decreased cell viability and inhibition of CIP2A, MYC and CDK1 protein expression. In AML patient samples, *ARPP19* mRNA expression followed the disease activity and was substantially lower at diagnosis in patients whose disease did not relapse during follow up. In addition, we discovered a novel CIP2A splicing variant NOCIVA that could act as a prognostic and predictive biomarker in AML and CML. The NOCIVA protein binds to PP2A-B56α, but whereas CIP2A mainly resides in the cytoplasm, NOCIVA translocated to the nucleus. AML and CML patient samples were found to overexpress *NOCIVA* mRNA. Elevated *NOCIVA* expression at diagnosis was identified as a biomarker of inferior overall survival in AML patients. High *NOCIVA* expression assessed at chronic phase CML diagnosis also associated with adverse event free survival exclusively in imatinib treated patients.

KEYWORDS: cancer, PP2A, biomarker, ARPP19, CIP2A, NOCIVA, OCT4, AML, CML, HNSCC, cancer stem cells, alternative splicing

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Biolääketieteen laitos
Patologia
Turun biotiedekeskus
ELEONORA MÄKELÄ: Novel PP2A biomarkers in cancer
Väitöskirja, 210 s.
Molekyylilääketieteen tohtoriohjelma
Maaliskuu 2021

TIIVISTELMÄ

Jotta ihmisen solu voi muuttua syöpäsoluksi, kasvunestäjäproteiini Proteiini-fosfataasi 2A:n (PP2A) toiminta tulee olla estynyt. Useimmiten PP2A:n toiminta on syövissä estynyt PP2A-inhibiittoriproteiinien (PAIP), kuten CIP2A, ARPP19, SET, PME-1 ja TIPRL, yli-ilmentymisestä johtuen. Tässä väitöstutkimuksessa tutkittiin edellä mainittujen PAIP:n toimintaa akuutissa ja kroonisessa myeloidisessa leukemiassa (AML ja CML) sekä pään ja kaulan alueen syövässä (HNSCC, engl. head and neck squamous cell carcinoma). Samalla tarkasteltiin PAIP:n kykyä toimia syöpäbiomarkkereina kyseisissä syövissä.

HNSCC:n yksi tärkeimmistä hoitomuodoista on sädehoito. Biomarkkereita, jotka ennustavat HNSCC kasvaimen sädeherkkyyttä ja jotka voisivat toimia kohdemolekyyleinä lääkkeille kasvaimen sädeherkkyyden parantamisessa, etsitään yhä. Väitöskirjani osatyössä havaitsimme OCT4:n säätelevän CIP2A:n ilmentymistä HNSCC-soluissa ja osoitimme kyseisen säätelyn edesauttavan HNSCC-solujen säderesistenssin syntyä. Kyseisiä markkereita voitaisiin tulevaisuudessa käyttää HNSCC kasvainten säderesistenssi arvioimiseen ja sädeherkkyyden parantamiseen.

Genetiikkaa myeloidisten leukemioiden takana on tutkittu lähivuosina paljon, mutta mekanismeja, jotka edesauttavat leukemian kehitystä ja uusiutumista ei vielä täysin tunneta. Väitöskirjani toisessa osatyössä havaitsimme AML potilaan diagnoosihetken ARPP19-mRNA ilmentymistason ennustavan taudin uusiutumista. ARPP19-proteiinin osoitettiin säätelevän MYC, CDK1 ja CIP2A onkoproteiinien ilmentymistä ja edistävän AML-solujen elinkykyisyyttä. Lisäksi kolmannessa osatyössä löydettiin täysin uusi CIP2A-mRNA variantin (NOCIVA), joka muodostuu vaihtoehtoisella silmukoinnilla ja joka voisi toimia ennusteellisena biomarkkerina sekä AML:ssä että CML:ssä. NOCIVA-proteiini pystyy sitoutumaan PP2A-B56α, mutta toisin kuin CIP2A, se ilmentyy pääasiallisesti tumassa. Sekä AML että CML potilasnäytteissä NOCIVA oli yli-ilmentynyt. AML:ssä korkean *NOCIVA* ekspression havaittiin korreloivan potilaan huonon ennusteen kanssa ja CML:ssä korkea *NOCIVA* ekspressio assosioitui lyhyemmän event free survival:n kanssa vain potilailla, jotka oli hoidettu imatinibillä.

AVAINSANAT: syöpä, PP2A, biomarkkeri, ARPP19, CIP2A, NOCIVA, OCT4, AML, CML, pään ja kaulan alueen syöpä, syövän kantasolu, vaihtoehtoinen silmukointi

Table of Contents

Abbreviations

AA	Amino acid
ABL	ABL proto-oncogene 1, non-receptor tyrosine kinase
AML	Acute myeloid leukemia
AS	Alternative splicing
ARPP19	cAMP-regulated phosphoprotein 19
BM	Bone marrow
CDK	Cyclin-dependent kinase
CIP2A	Cancerous inhibitor of protein phosphatase 2A
CML	Chronic myeloid leukemia
CMR	Complete molecular response
CP	Chronic phase
CR	Complete remission
CSC	Cancer stem cell
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
EFS	Event free survival
ELN	European LeukemiaNet
ENSA	Alpha-endosulfine
ESS	Exonic splicing silencer
EVI1	Ectopic viral integration site 1
FAB	French-American-British classification
FBS	Fetal bovine serum
FFP	Freedom from progression
FLT3	Fms related receptor tyrosine kinase 3
FLT3-ITD	FLT3 internal tandem duplication
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
GST	Glutathione S-transferase
GWL	Great wall kinase
HNC	Head and neck cancer
HNSCC	Human head and neck squamous cell carcinoma

HPV	Human papilloma virus
HR	Hazard ratio
HSC	Hematopoietic stem cell
HSCT	Hematopoietic stem cell transplantation
LSC	Leukemia stem cell
MDS	Myelodysplastic syndrome
MYC	C-MYC proto-oncogene
MRD	Minimal/Measurable residual disease
mRNA	Messenger ribonucleic acid
NGS	Next generation sequencing
NHEK	Normal epidermal keratinocyte
NOCIVA	Novel CIP2A variant
NPM1	Nucleophosmin 1
OCT4	Octamer-binding transcription factor 4
OS	Overall survival
PAIP	PP2A inhibitor protein
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PME-1	Protein phosphatase methyl esterase 1
PP2A	Protein phosphatase 2A
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RT	Room temperature
RQ-PCR	Quantitative real-time polymerase chain reaction
shRNA	Small hairpin RNA
siRNA	Small interfering RNA
SET	SET nuclear proto-oncogene
SV40	Simian virus 40
TCGA	The Cancer Genome Atlas
TIPRL	TOR signaling pathway regulator
TKI	Tyrosine kinase inhibitor
TRF	Treatment-free remission
UTR	Untranslated region
WHO	World health organization
WT1	Wilms' tumor 1
2G TKI	Second generation tyrosine kinase inhibitor
3G TKI	Third generation tyrosine kinase inhibitor

List of Original Publications

This thesis by Eleonora Mäkelä (née Sittig) is based on the following original publications, which are referred to in the text by their Roman numerals:

I Ventelä S., Sittig E., Mannermaa L., Mäkelä J-A., Kulmala J., Löyttyniemi E., Strauss L., Cárpen O., Toppari J., Grénman R., Westermarck J. CIP2A is an OCT4 target gene involved in head and neck squamous cell cancer radioresistance. *Oncotarget,* 2015; 6(1): 144-58.

II Mäkelä E., Löyttyniemi E., Salmenniemi U., Kauko O., Varila T., Kairisto V., Itälä-Remes M., Westermarck J. Arpp19 promotes Myc and Cip2a expression and associates with patient relapse in acute myeloid leukemia. *Cancers (Basel)*, 2019; 11(11): 1774.

III Mäkelä E., Pavic K., Varila T., Salmenniemi U., Löyttyniemi E., Nagelli S., Kähäri V-M., Clark R. E., Bachanaboyina V. K., Lucas C. M., Itälä-Remes M., Westermarck J. Discovery of NOvel CIP2A VAriant (NOCIVA) and its clinical relevance in myeloid leukemias. Accepted for publication in *Clinical Cancer Research*.

1 Introduction

Cancer is among the leading causes of death and a major public health issue worldwide. The word "cancer" stands for multiple different cancer diseases that vary substantially in their clinical picture and prognosis, yet also share molecular and phenotypical features. Cancer affects millions of people and in the year 2020 about 18 million new cancer cases and nearly 10 million cancer related deaths globally have occurred. Although the incidence of cancer is increasing worldwide mainly due to changes in lifestyle and lengthened lifespan, cancer mortality has at the same time declined due to better treatment options and early diagnosis. Discoveries in cancer research are the basis for cancer biomarker and drug development that has already led, but will also continue to lead, to better patient stratification and ultimately superior cancer patient care.

Protein phosphorylation, the most common post-translational modification of proteins which involves the attachment or removal of phosphatase groups, represents a fast and reversible way to regulate protein function. Phosphorylation-dependent signal transduction is a highly controlled process that requires coordinated and timely regulation of both protein kinases and phosphatases and deregulation in this balance underlies many human diseases, including cancer. Inactivation of a major serine/threonine phosphatase, protein phosphatase 2A (PP2A), acts as one of the prerequisites for the malignant transformation of a human cell. In human cancer, PP2A is in majority of cases inactivated by the elevated expression of endogenous PP2A inhibitory proteins, such as SET, ARPP19, PME-1 and CIP2A.

Although many aspects of PP2A signaling have been extensively studied, the role of PP2A inhibitor proteins as clinically useful biomarkers remains largely unexplored. In addition, CIP2A has been demonstrated to be a prominent oncoprotein in multiple different cancers, yet the possible mRNA and protein variant forms of CIP2A remain elusive. This thesis focuses on PP2A inhibitor proteins as cancer biomarkers in acute and chronic myeloid leukemias and in head and neck squamous cell carcinoma.

2 Review of the Literature

2.1 Cancer

Cancer is the name given to a collection of more than 100 distinct, but related diseases characterized by the uncontrolled growth of abnormal cells that have acquired the ability to invade nearby or distant tissues in the body (Hanahan and Weinberg 2000). Cancer has a major impact on society across the world as it is globally an important cause of mortality and morbidity (Bray et al. 2018). According to GLOBOCAN 2018 global cancer statistics, there were about 18.1 million new cancer cases diagnosed and 9.6 million cancer related deaths worldwide in 2018 (Bray et al. 2018; Ferlay et al. 2019). According to the world health organization (WHO) estimates in 2015, cancer is the first or second leading cause of deaths worldwide.

Cancer incidence and mortality are rapidly growing worldwide and the number of new cancer cases per year is expected to rise to 29.5 million by 2040 (Bray et al. 2018). The reasons for the cancer burden increase are complex but mirror both growth and aging of the population, as well as changes in the lifestyle behaviors associated with higher cancer risk, several of which are connected with socioeconomic development. Importantly, alone in the United States, overall national expenditures for cancer care in 2015 were $183 billion and the costs are estimated to increase to $246 billion by 2030, based only on population growth (Mariotto et al. 2020). In the future years, cancer-attributed medical care costs are additionally likely to rise worldwide as the cancer prevalence increases, but also due to the rising burden of cancer care among cancer survivors. Novel and usually more expensive treatments are also adopted as standards of care, and this will additionally have an impact on the cancer related expenditures. Thus, cancer-attributable cost estimates by cancer site, stage and disease phase are key inputs for cost effectiveness analyses for cancer care. These analyses combined with the data from molecular cancer studies then enable the formation of improved patient stratification strategies that in the end will benefit both the patient and the society.

2.1.1 Malignant transformation

During malignant transformation, cells need to acquire features that enable them to evade from normal cellular control mechanisms. The most distinguishing property of transformed cells from normal cells is the ability to escape from growth control and to divide in an uncontrolled manner. In normal cells, a subtle balance exists between tumor suppressors, which restrain cellular growth, and proto-oncogenes, which promote cellular growth and proliferation. During the multistep process of carcinogenesis, cells obtain several genetic and epigenetic alterations both in tumor suppressor genes and proto-oncogenes, leading to dysregulation of normal cellular functioning, and eventually to formation of a continuously growing cellular mass (Hahn and Weinberg 2002; Vogelstein et al. 2013). The sequence, frequency, and the type of the alterations is dependent on the cancer type and may differ between individual cancers types. Moreover, random mutations in the genome are common and can in most cases be repaired before they cause deleterious effects. However, alterations that provide selective growth advantage to the cells, and thus drive carcinogenesis, are sustained and start to accumulate over a span of cellular divisions. More than 138 cancer driver genes, participating in 12 different signaling pathways that regulate proliferation, cell survival, differentiation, DNA repair and other processes involved in genome integrity, have been recognized (Vogelstein et al. 2013). Intriguingly, a very recent study with more than 28 000 tumors from 66 cancer types identified 568 cancer genes with the ability to drive tumorigenesis (Martínez-Jiménez et al. 2020), emphasizing the need for continuous refinement of our understanding of the genetic basis of cancer.

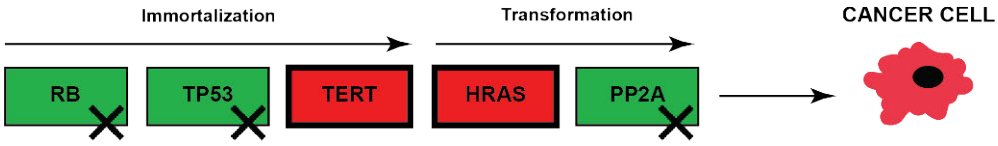

Figure 1. Minimal genetic alterations needed for a malignant transformation of a human fibroblast. Inactivated tumor suppressors are presented in green and activated oncogenes in red. Active telomerase enzyme TERT alongside with inhibited tumor suppressors TP53 and RB are required for the immortalization of a human cell. The inhibition of PP2A activity with an active oncogenic H-RAS are the last prerequisites for the complete human cell transformation into cancer cell.

The minimum genetic requirements for the cellular transformation of cultured human fibroblasts into cancer cells were identified almost two decades ago (Hahn et al. 1999; Yu, Boyapati, and Rundell 2001; Hahn et al. 2002). The experimental transformation of human cells was proven to be more difficult than transformation of mouse cells, and the fundamental difference was identified in the specific need

for protein phosphatase 2A (PP2A) inactivation along with the constitutive activation of telomerase (TERT) in human cells (Chang et al. 1985; Rangarajan et al. 2004). Whereas only two genetic alterations were needed for transformation of mouse fibroblasts, namely an activating mutation in HRAS and inactivation of either retinoblastoma (RB) or TP53 (p53), five alterations were needed to fully transform human fibroblasts (Figure 1). It was demonstrated that in human fibroblasts three alterations are required for cellular immortalization (Figure 1): the expression of active telomerase (TERT) and the inhibition of tumor suppressor proteins TP53 and RB (Hahn et al. 1999; Hahn and Weinberg 2002). However, the complete transformation of human cells was accomplished only after expression of activated HRAS oncogene and addition of the simian virus 40 (SV40) small t-antigen (ST) on the immortalized cells (Yu, Boyapati, and Rundell 2001; Hahn et al. 2002). Owing to SV40 ST's ability to inactivate tumor suppressor PP2A via interaction with A- and C-subunits of the PP2A complex (Pallas et al. 1990), PP2A inhibition was established as one of the prerequisites for malignant human cell transformation. However, nowadays it's known that what holds true for human fibroblast transformation, does not apply to all human cell types. Cell type specific differences exist in the requirements for malignant transformation, such as for example the requirement of activation of H-RAS downstream effectors guanine nucleotide exchange factors (GEFs), Raf or PI3K in some human cell types (Rangarajan et al. 2004).

2.1.2 Hallmarks of cancer

The development of malignant tumors from the priming population of transformed cells is a multistep process, in which the malignant cells need to obtain characteristics that enable escape from normal cellular limitations. In 2000 Hanahan and Weinberg defined the "Hallmarks of cancer", the six key alterations in cellular signaling and behavior, that enable human cancer development, and are considered as commonalities among all different cancers (Figure 2). These hallmarks include avoiding programmed cell death, evading signaling leading to growth suppression, sustaining proliferative signaling without the need of extracellular stimulus, enabling replicative immortality, inducing angiogenesis, and activation of invasion and formation of metastasis (Hanahan and Weinberg 2000). These crucial attributes allow initiation, promotion, malignant transformation and progression of cancer during carcinogenesis.

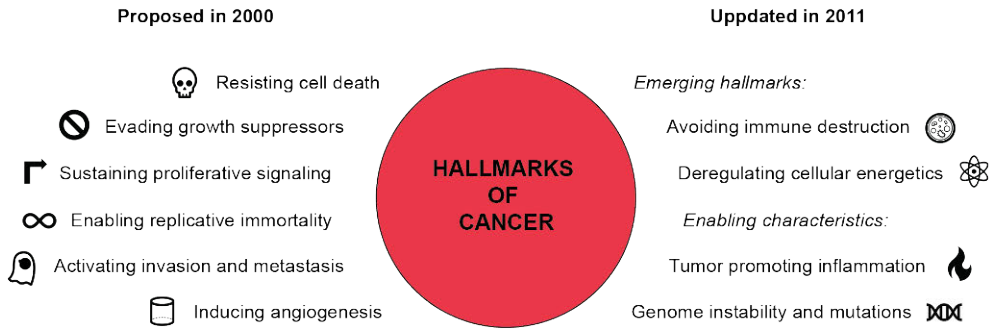

Figure 2. Hallmarks of cancer by Hanahan and Weinberg.

In 2011, Hanahan and Weinberg revised the hallmarks and proposed two emerging additions (Figure 2): avoiding immune destruction and surveillance, and deregulating cellular energy metabolism (Hanahan and Weinberg 2011). At the same time cancer enabling characteristics were proposed. Genomic instability and mutations form the basis for all the hallmarks of cancer by introducing the alterations in the genetic material, whereas the tumor promoting inflammation enables cancer progression and sustenance. Hence, cancer is no longer considered as a disease caused by genetic alterations alone, but as a disease of a complex cascade of events ranging from mutations in the genome to changes in systemic features. Furthermore, it has become evident that variation in growth, apoptosis, and other hallmarks of cancer properties of the individual cells within a tumor or cancer *per se*, are far from homogeneous. Thus, cancer cannot either be considered as a mass of homogeneous cells with equal attributes, but instead, as complex networks where individual cells exhibit various properties but at the same time they function together to support the growth and maintenance of a cancer as a whole.

2.1.3 Head and neck squamous cell carcinoma (HNSCC)

In 2020, head and neck cancer (HNC) is expected to affect approximately 888 000 new patients worldwide, and around 50% of the patients will succumb to their disease (Ferlay et al. 2013; Bray et al. 2018). This makes HNC the sixth most common type of cancer by incidence worldwide, and it is also the leading cause of cancer death among men in India and Sri Lanka (Bray et al. 2018). HNC is a heterogeneous collection of diseases that include cancers in the mouth, lips, larynx, nose, throat and salivary glands, but most of the HNCs originate from the epithelial linings of the upper aero-digestive tract (oral cavity, larynx, oropharynx and hypopharynx), and about 90% of the HNCs are squamous cell carcinomas (HNSCCs)(Chow 2020). The main environmental and lifestyle risk factors contributing to HNSCC development are tobacco smoking and heavy alcohol

consumption. In addition, human papilloma virus (HPV) infection is associated with the generation of a subclass of HNSCC that form a separate clinicopathological and molecular HNSCC group with superior prognosis. Although HNSCC tumors have remarkable genetic heterogeneity and it has even been shown that HPV-positive and HPV-negative HNSCCs have different genetic drivers (Lechner et al. 2013), a common nominator in the HNSCC pathogenesis is the formation of genetically altered mucosal preneoplastic fields from which the carcinomas arise. Perturbation of all the major pathways, TP53, RB, PI3K/AKT/mTOR, EGFR and NOTCH, have been documented in HNSCC (Ausoni et al. 2016).

Whereas older patients' HNSCC associated with heavy use of tobacco and alcohol are slowly declining globally, cases of HPV-associated HNSCC are increasing mainly among younger people in northern Europe and North America (Chow 2020). The classification of HNSCCs into disease stages is done by using the globally recognized tumor, node, metastasis (TNM) staging system along with the Union for International Cancer Control (UICC) system (Amin et al. 2017). Multiple promising biomarkers have been identified in HNSCC, but none of the reported biomarkers have yet reached clinical use (Hsieh et al. 2019). This means that there are currently either no validated biomarkers available for prediction of therapy response of HNSCC. Thus, treatment decisions for HNSCC patients are determined by taking into consideration factors such as patient's general health, and the anatomical site, stage and surgical accessibility of the tumor. As the current data are insufficient to recommend changes in treatment or less-intensive treatment for HPV-associated disease, HPV-negative and HPV-positive tumors are for the time being treated similarly (Chow 2020).

2.1.3.1 Radiation resistance in HNSCC

Radiation is one of the key treatment modalities for the management of HNSCC, and therefore tumor's sensitivity to radiotherapy often determines local control of HNSCC (Hutchinson, Mierzwa, and D'Silva 2020). Treatment of early stage HNSCC involves single-modality therapy with either surgery or radiation, whereas in order to achieve a better local control of an advanced HNSCC tumor, combined approaches including surgery followed by adjuvant therapy or definitive chemoradiation are conducted (Furness et al. 2011). Apart from the effects of radiotherapy through direct DNA damage and indirect damage from free radical formation ultimately leading to apoptosis of the cancer cells, radiation additionally induces antitumor immune responses that are important in the indirect tumor cell killing (Weichselbaum et al. 2017). However, one of the generally acknowledged challenge in HNSCC treatment is the high radiation resistance of the HNSCC cells.

Multiple different approaches have been taken to clarify the causes of intrinsic HNSCC radioresistance. Major causes identified include mutations and/or aberrant expression of proteins in cellular pathways involved in DNA damage repair, phenotypic and functional heterogeneity of HNSCC including hypoxic areas within HNSCC tumors, and the presence of cancer stem cells (CSCs) in the HNSCC tumors. Preclinical and clinical evidence have additionally demonstrated that alterations and/or distribution of three main pathways, EGFR, PI3K/AKT/mTOR and TP53 signaling cascades, play a crucial role in radioresistance development in HNSCC. For example, EGFR is overexpressed in over 90% of HNSCC (Harari, Wheeler, and Grandis 2009), disruptive and nondisruptive TP53 mutations have been detected in up to 85% HNSCC (Stransky et al. 2011) and activating mutations in the PIK3CA gene are found in 30% of HNSCC (Lui et al. 2013). All these alterations help the cancer cells to evade from the radiation-induced apoptosis and to enhance DNA damage repair. Furthermore, aberrant expression of multiple proteins, such as TRIP13 (Banerjee et al. 2014), Ku80 (Moeller et al. 2011) and ATM (Mansour et al. 2013), involved in nonhomologous end joining (NHEJ) that is the major pathway responsible for repairing radiation-induced DNA double-strand breaks, have been reported to contribute to radioresistance in HNSCC. In addition, features linked to CSCs, such as pluripotency and self-renewal have been proposed to be one explanation for the radiation resistance observed in HNSCC. Intriguingly, there are even studies reporting that radiation itself could induce non-CSCs to become CSCs (Ghisolfi et al. 2012). Although the underlying mechanism of CSC mediated radioresistance remain to be elucidated, it has been shown that these slow cycling cells display elevated efficiency in DNA damage repair, enhanced scavenging of reactive oxygen species, elevated cell survival and reduced apoptosis induction (Peitzsch et al. 2019).

Identification of new targetable players in HNSCC radioresistance could provide novel opportunities for radiosensitation and further lead to combination therapies with improved disease control and patient survival.

2.1.4 Acute myeloid leukemia (AML)

Acute myeloid leukemia (AML) is the most common acute leukemia affecting adults with around 20 000 new cases diagnosed in the United States (Siegel, Miller, and Jemal 2020) and about 200 in Finland per year (https://syoparekisteri.fi/tilastot/tauti-tilastot/). The incidence of AML increases with age and the median age at diagnosis in the United States is 68 years (Siegel, Miller, and Jemal 2020). AML is a heterogenous hematological malignancy that arises from poorly differentiated cells in the myeloid lineage and leads to disruption of normal hematopoiesis. AML emerges from the hematopoietic progenitor cells that have obtained a differentiation

block, increased proliferation and resistance to apoptosis mainly due to genetic alterations. As the substitution of normal blood cells with leukemic blasts in the bone marrow causes recurrent infections, anemia and bleeding that are typically fatal within weeks or months if left untreated, AML accounts for one of the most aggressive cancer types with only 28% 5-years survival (Döhner, Weisdorf, and Bloomfield 2015).

The only proven lifestyle-related risk factor for developing AML is tobacco smoking. Other risk factors for AML include other blood disorders, chemical exposures, ionizing radiation, previous treatment with chemotherapies, and genetics (Döhner et al. 2017). Some people with AML have one or more known risk factors, but many have none, thus the most likely explanation for AML genesis lies on the genetic alterations accumulated over a person's lifetime. Genomic changes accumulate over time, which could partly explain why AML appears more frequently in older adults. However, it has been estimated that on average AML patients harbor 13 mutations, which is actually less than in most adult cancers (Ley et al. 2013). Nevertheless, AML is a highly heterogeneous collection of blood cancers with acknowledged mutational and cytogenetic complexity.

2.1.4.1 Classification of AML

To look for leukemia, blood tests are generally the first tests conducted. AML diagnosis is established when at least 20% of the cells in the peripheral blood or bone marrow are defined as blasts based on morphological examination (Arber et al. 2016; Döhner et al. 2017). In addition to morphological evaluation, an extensive set of laboratory tests, such as complete blood count and differential count, immunophenotyping, cytogenetics analyses, and screening for mutations and gene rearrangements, are incorporated in order to confirm and specify an AML diagnosis (Döhner et al. 2017). During the diagnosis process, bone marrow samples are taken for the evaluation of morphology, lineage and immunophenotype of the leukemic cells as well as to provide material for cytogenetic and genetic analyses. The differentiation of AML from other forms of leukemia, in particular chronic myeloid leukemia (CML) and acute lymphocytic leukemia (ALL), is also important in terms of therapeutic implications.

The two most commonly used classification schemes for AML subtypes are the older French-American-British (FAB) system and the newer World Health Organization (WHO) system. In the FAB system, AML is divided into eight different subtypes, M0-M7, based on the cell type and the differentiation stage of the leukemic cells (Table 1) (Walter et al. 2013). In M0 and M1 subtypes the block in differentiation has occurred at a very early stage of myeloid progenitor cell development. In M4 AML the leukemic cells show already myelomonocytic

differentiation and in the M5 subtype monocytic differentiation. In M6 AML the disease is mainly orchestrated by immature erythroid precursors and in M7 AML by platelet precursors called megakaryoblasts. The FAB classification has been used for decades, but was mostly replaced by the WHO classification (Arber et al. 2016) in 2001 as the WHO classification offers a better prognostic value and takes heed of prominent mutations and cytogenetic alterations identified in AML.

Table 1. The French-American-British (FAB) classification in AML. Adapted from (Walter et al. 2013).

FAB subtype	Name	Prevalence
M0	Undifferentiated acute myeloblastic leukemia	5%
M1	Acute myeloblastic leukemia with minimal maturation	20%
M2	Acute myeloblastic leukemia with maturation	25%
M3	Acute promyelocytic leukemia (APL)	10%
M4	Acute myelomonocytic leukemia	20%
M4 eos	Acute myelomonocytic leukemia with eosinophilia	5%
M5	Acute monocytic leukemia	10%
M6	Acute erythroid leukemia	4%
M7	Acute megakaryoblastic leukemia	1%

2.1.4.2 Prognostic factors and prognosis of AML

Genetic abnormalities are powerful prognostic factors in AML. Approximately half of the patients with AML harbor chromosomal abnormalities (Papaemmanuil et al. 2016) and chromosomal translocations that cause the formation of fusion genes, such as *CBFB-MYH11, RUNX1-RUNX1T1, PML-RARA* and *MLL,* have been known in AML for several decades already (Mrózek et al. 1997). These fusion genes frequently associate with the FAB subtypes and are significant determinants of treatment response (Table 2). However, next generation sequencing (NGS) studies within the past 10 years have only started to reveal the mutational heterogeneity and complexity presented in AML. Based on the AML genome sequencing data (Ley et al. 2013; Papaemmanuil et al. 2016; Grimwade, Ivey, and Huntly 2016; Tyner et al. 2018), the most commonly mutated genes in AML involve genes associated with DNA methylation, such as *DNMT3A* in 26% and *IDH1/2* in 20% of cases, and with activated signaling cascades, such as *FLT3-ITD/TKD* in 28% and *NRAS/KRAS* in 14% of cases. Additionally, nucleophosmin (*NPM1*) has been reported to be mutated

in 27% of AML cases. Although it has been illustrated that the co-occurrence of these mutations can have a prognostic impact in overall survival (Papaemmanuil et al. 2016), the prognostic impact of many of the AML markers is context dependent and the effect of any given alteration depends on the presence or/and absence of another. The mutational status of AML patients has also been associated with response to drugs and instances of drug sensitivity that are specific to combinatorial mutational events have also been reported (Tyner et al. 2018). Thus, the future challenge and hope is in precision oncology that takes into consideration the whole spectrum of genetic lesions of an individual to finetune the risk stratification and treatment of a single patient throughout the course of AML patient care (Gerstung et al. 2017).

With the routine use of molecular diagnostics, the identification of recurrent genetic lesions has become a general practice and the core in determining prognosis for AML patients. The current risk classification of the European LeukemiaNet (ELN, latest update in 2017 (Döhner et al. 2017)), which is based on the WHO classification, divides patients into three distinct risk groups: favorable, intermediate and adverse (Table 2). The combined data from cytogenetic and molecular analyses provide the strongest prognostic information available, aiming to predict outcome of both remission induction and post-remission therapy. However, the studies that led to the ELN-2017 recommendations involved AML patients receiving traditional cytarabine-based induction chemotherapy, but the treatment with alternative first-line approaches such as FLT3 inhibitors, hypomethylating agents, venetoclax-based regimens or immunotherapeutics, may alter the prognosis of these patients and therefore induce changes in the genetic risk classification in the near future (Herold et al. 2020). Furthermore, predicting long-term patient outcomes based solely on the pretreatment genetic characteristics is way too inaccurate. Thus, sequential analyses of measurable residual disease (MRD) during and after treatment by real-time quantitative PCR (RQ-PCR), digital PCR, NGS-based technologies and flow cytometry have emerged as indispensable tools to assess response to therapy and to refine prognosis.

Table 2. Risk stratification by molecular profiling based on ELN-2017 classification.

Risk category	Genetic abnormality
FAVORABLE	t(8;21)(q22;q22.1); *RUNX1-RUNX1T1* inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11* Mutated NPM1 without FLT3-ITD or with FLT3-ITD^{low} Biallelic mutated CEBPA
INTERMEDIATE	Mutated *NPM1* and FLT3-ITD^{high} Wild-type *NPM1* without FLT3-ITD or with FLT3-ITD^{low} (without adverse-risk genetic lesions) t(9;11)(p21.3;q23.3); *MLLT3-KMT2A* Cytogenetic abnormalities not classified as favorable or adverse
ADVERSE	t(6;9)(p23;q34.1); *DEK-NUP214* t(v;11q23.3); KMT2A rearranged t(9;22)(q34.1;q11.2); *BCR-ABL1* inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM(EVI1) −5 or del(5q); −7; −17/abn(17p) Complex karyotype,§ monosomal karyotype Wild-type NPM1 and FLT3-ITD^{high} Mutated RUNX1¶ Mutated ASXL1¶ Mutated TP53

§ Three or more chromosome abnormalities but not including recurrent translocations.
¶ These markers should not be used if they co-occur with favorable risk AML subtypes.

The core of therapy for fit AML patients has not significantly changed in recent years. Upon diagnosis, first evaluation assesses whether a patient is suitable for intensive induction chemotherapy. Induction therapy for AML has remained unchanged for the last 50 years and the standard treatment involves three days of an anthracycline (idarubicin or daunorubicin) and seven days of cytarabine aiming for complete remission (CR). With this approach, up to 85% of the patients under the age of 60 achieve CR, but AML will almost certainly recur if no further treatment is given after a CR (Döhner, Weisdorf, and Bloomfield 2015). CR in AML is defined as blasts less than 5% in the bone marrow with the absence of circulating blasts, blasts with Auer rods and extramedullary disease (Döhner et al. 2017). If CR is reached, the patients belonging to intermediate and adverse risk groups receive an allogeneic hematopoietic stem cell transplantation (HSCT), whereas patients in the favorable risk group can be further treated with intensive chemotherapy. These consolidation therapy schemes can result in 30 to 40% curative rates in AML patients under 60 years of age (Petersdorf et al. 2013). However, the outlook for AML remains unsatisfactory as up to 50% of patients will relapse and the prognosis for the relapsed or chemorefractory patients remains dismal. Also, many AML patients are

unfit for conventional therapy because of age or co-morbidities. Standard of care for these patients is therefore significantly different and consists primarily of hypomethylating agents (HMAs) azacitidine or decitabine or low dose cytarabine (Döhner et al. 2017), but more recently HMA plus venetoclax (Agarwal et al. 2021). The treatment landscape for these AML patients as well as relapsed patients has rapidly changed during the past 5 years with several new targeted drugs approved. Nevertheless, a plateau has been reached with the current treatments and novel targeted therapies that could improve the prognosis, as well as novel markers for better MRD monitoring, are urgently needed for patients with AML.

2.1.4.3 Alternative splicing in AML

Alternative splicing (AS) is a normal physiological phenomenon that highly increases the diversity of mRNAs in the eukaryotic transcriptome. Whereas in constitutive splicing the introns are removed, and the exons ligated in the order in which they appear in a gene to form a mature messenger RNA (mRNA), AS deviates from this preferred sequence resulting in several altered forms of mRNAs from the same gene. As up to 95% of genes that have multiple exons are alternatively spliced (Pan et al. 2008), AS also ensures better environmental fit via increased protein diversity. Multiple modes of AS have been identified, but the molecular mechanisms of AS are highly variable, and new mechanisms are still constantly being found. Five main modes of AS that are generally recognized are the intron retention, mutually exclusive exons, exon skipping, alternative donor site and alternative acceptor site (Wang et al. 2015). The exons that are included in the mature mRNA are mainly defined by the interaction between cis-acting elements and trans-acting binding factors. These cis-acting elements include exonic and intronic splicing enhancers (ESEs and ISEs) that are bound by splicing promoting factors, such as serine/arginine-rich family of nuclear phosphoproteins (SR-proteins), whereas exonic and intronic splicing silencers (ESSs and ISSs) interact with splicing inhibiting factors, such as heterogeneous nuclear ribonucleoproteins (hnRNPs) (Wang and Burge 2008). The combination of these elements on a gene then results in the promotion or inhibition of spliceosome assembly on a specific splice site.

Aberrant AS has been denoted as one of the underlying causes for cancer development (Sveen et al. 2016). Disturbance of AS by epigenetic modifications, spliceosome gene mutations, or alterations of the cellular composition, localization or activity of the splicing factors lead to aberrant AS. Recent pan-cancer studies have indicated that there are on average 20% more AS events in tumors than in healthy samples (Climente-González et al. 2017; Kahles et al. 2018). As AS is an essential part of normal hematopoiesis and a necessity for proper cellular differentiation

(Grech et al. 2014), the role of abnormal AS in hematological malignancies has gained understandable attention. Multiple studies have underlined recurrent splicing factor mutations as important drivers of various hematological cancers (Lindsley et al. 2015; Papaemmanuil et al. 2016; Taskesen et al. 2014). Interestingly, current evidence proposes a central role for AS abnormalities especially in leukemia pathogenesis (Yang et al. 2018; Adamia et al. 2014). In AML, recurrent mutations in spliceosome machinery and splicing factors (*U2AF1, SRSF2* and *SF3B1*) as well the genome wide aberrant splicing events have been identified as a distinctive component of the disease (de Necochea-Campion et al. 2016).

2.1.5 Chronic myeloid leukemia (CML)

Chronic myeloid leukemia (CML) is a clonal myeloproliferative disorder that is characterized by the Philadelphia (Ph) chromosome and driven by its product, the constitutively active BCR-ABL1 tyrosine kinase (Faderl et al. 1999; Lugo et al. 1990). Unlike in other myeloid malignancies, *BCR-ABL1* is usually the only genetic alteration detected in newly diagnosed CML patients. If any co-occurring mutations are detected, most are still of unclear significance (Branford et al. 2019). As in case of many other cancers, the incidence of CML increases with age and the median age at diagnosis in the United States is 65 years, with about 8 500 new cases diagnosed per year (https://seer.cancer.gov).

CML is divided into three distinct phases: chronic, accelerated, and blast, that help to plan treatment and predict prognosis. In chronic phase (CP) the blood and bone marrow contain less than 10% blasts, whereas in the blast phase the blast count is 20% or more and the blasts often resemble the immature cells observed in patients with other types of leukemia, specifically AML for most patients (Faderl et al. 1999). Importantly, CML is usually easily diagnosed as the CP leukemic cells of more than 95% of patients have the Ph chromosome and the presence of the 9;22 translocation that accounts for the fused *BCR-ABL1* gene can be easily detected from a peripheral blood sample by RQ-PCR (Hochhaus, Baccarani, et al. 2020). About 90% of patients have CP CML at the time of diagnosis, but without effective treatment, CML will move into accelerated phase and further into blast phase in about three to four years after diagnosis (Faderl et al. 1999). However, the pace of CML treatment research has been exceptionally rapid, with several drug approvals and major changes in the treatment strategies in the past two decades. While the median survival in CML used to be 4 to 6 years, with the advent of the highly effective targeted therapies, tyrosine kinase inhibitors (TKIs), a survival rate that is equivalent to that in the general population is expected for most patients with CML in the 2020s (Hochhaus, Baccarani, et al. 2020).

Imatinib, a first generation TKI introduced in 2001, remains the most cost-effective and standard first-line therapy for patients with CP CML, and has led to outstanding treatment results (Gambacorti-Passerini et al. 2011; Hochhaus et al. 2017). Imatinib is often considered as a model for targeted cancer therapy due to its outstanding ability to reduce CML disease progression as well as CML related deaths. At the moment, five TKIs are approved for the treatment of CML, including a first generation TKI imatinib, second generation (2G) TKIs dasatinib, nilotinib, and bosutinib, and a third generation (3G) TKI ponatinib. The choice of TKI therapies for individual patients is decided by considerations of efficacy, early and late toxicity, tolerability, and drug costs (Hochhaus, Baccarani, et al. 2020). Currently, the hematologic toxicity (causing neutropenia, thrombocytopenia or anemia) of the different TKIs is considered equal (Steegmann et al. 2016), but the side-effects that affect tolerability and quality of life are demonstrated to be broader for the 2G and 3G TKIs than for imatinib. These adverse side-effects also cause a treatment change in a large number of patients. It has become evident that achieving a stable deep molecular response (DMR), defined as *BCR-ABL1* transcript level on the international scale $\leq 0.01\%$, at high rates and more quickly is an important treatment goal in CML patient care. Achievement of DMR within 12 months after the initiation of treatment for newly diagnosed CP CML has been associated with a very low risk of long-term progression (Hochhaus, Baccarani, et al. 2020). Furthermore, whereas CP CML is solely driven by the *BCR-ABL1* gene product, the advanced phases of CML are directed by multiple pathways and complex genomic alterations, thus the importance of initiating treatment early in the course of the disease has been underscored. Very recently, an important development in CML management has been the realization that TKI therapy can be successfully discontinued in some patients in so called treatment-free remission (TFR) (Hochhaus, Baccarani, et al. 2020). Stable DMR has been established as a prerequisite for making TFR feasible for potential cure. Although currently the majority of patients with CP CML receive imatinib as the first-line therapy, increasingly patients are directed to 2G TKI therapy as frontline treatment for the goal of higher and faster probabilities of a TFR attempt (Hochhaus, Breccia, et al. 2020; Braun, Eide, and Druker 2020). Identification of treatment strategies that maximize the possibility of TFR will certainly be the focus of future CML research.

TKIs and cost effectiveness

Despite advances mentioned above, long-term TKI therapy remains associated with decreased quality of life and increased financial burden. Because most patients with CML require lifelong treatment with TKI, TKI costs and cost effectiveness have

become cardinal issues for both patients and the society. At first hand, the 2G and 3G TKIs are much more expensive than imatinib, and this difference has further increased since the patent for imatinib expired in 2015 and generic imatinib compounds have become broadly available (Shih, Cortes, and Kantarjian 2019). Lifelong-cost estimates are thus nowadays key variables in the decision making for a frontline TKI and when changing to alternative medication in the course of treatment. The possibility of achieving a prolonged TFR should also be included in the lifelong-cost estimates, although at the moment this concerns only about 20% of the patients with CML (Hochhaus, Baccarani, et al. 2020). Multiple studies have been conducted to address these issues and the cost effectiveness of first-line treatments in various scenarios have been extensively examined (Padula et al. 2016; Shih, Cortes, and Kantarjian 2019; Yamamoto et al. 2019; Andrews and Lipton 2019). 2G TKIs as initial therapy that aim to increase the rate of sustained DMR and subsequently lead to increased rates of achieving TFR have also been included in these investigations. The European LeukemiaNet expert panel concluded from this data that in the present time when all the 2G (and 3G) TKIs are still under patent protection and as 80% of CML patients will never achieve a TFR, generic imatinib remains as the most cost effective initial treatment strategy for CP CML (Hochhaus, Baccarani, et al. 2020). Thus, until the 2G TKIs lose their patent protection, cost effectiveness will continue to be a significant issue when determining frontline, but also second-line TKI in CML patient care.

2.2 Cancer stem cells (CSCs)

Nowadays cancer is widely accepted to be a heterogeneous disease with high intratumor and intertumoral heterogeneity. Two main tumor propagation models have been proposed to explain the origin, maintenance, progression, and heterogeneity of cancer: the cancer stem cell (CSC) model, also known as the hierarchical model, and the stochastic model. Whereas the stochastic model is the historical one and based on any cell's ability to undergo significant number of genetic alterations for it to become cancerous, the stem cell model posits that nongenetic, mainly epigenetic determinants create hierarchically organized cancer tissues where a population of self-renewing CSCs lie at the apex and sustains the long-term clonal maintenance of the cancer. A common nominator between the models is the clonal evolution theory, which postulates that the expansion of a subpopulation of cancer cells, a clone, originates from an individual single cell (Figure 3). Essentially, the stochastic model suggests that all cells are able to be tumorigenic, making all tumor cells equipotent with the ability to differentiate or self-renew. Whereas, on the contrary, the CSC model postulates that only the CSCs may generate tumors through the stem cell processes of differentiation and self-

renewal. Furthermore, the CSC model hypothesizes that CSCs persist in tumors as a biologically distinct subpopulation and only they cause relapse and metastasis by giving rise to new tumors. Although previously considered as mutually exclusive models to describe cancer heterogeneity, it has been proposed that the stochastic and CSC models could be harmonized by integrating the role of genetic diversity and nongenetic influences in contributing to cancer heterogeneity in a unified model (Kreso and Dick 2014). (Kreso and Dick 2014; Cabrera, Hollingsworth, and Hurt 2015; Plaks, Kong, and Werb 2015; Batlle and Clevers 2017)

As unique subpopulations of cells emerge during tumorigenesis, various end results are possible: less fit subclones can be completely lost as the most fit subclone dominates, or multiple minor subpopulations can continue to exist alongside the dominant clone, acting as pool of cells from which evolution can continue (Figure 3). This is known as branching clonal evolution and evidence from multiple progressed cancers support this phenomenon in which multiple branches of subclones evolve independently in parallel (Batlle and Clevers 2017). In addition, emerging data from several studies in various cancers supports the concept of cancer plasticity, in which the microenvironment plays a crucial role in shaping the cancer cells phenotype by instructing reversible cellular state transitions that influence equally all the cells in the tumor, including the CSCs. It has been further postulated that CSCs exhibit plasticity by reversibly transitioning between stem cell and non-stem cell states (Cabrera, Hollingsworth, and Hurt 2015; Batlle and Clevers 2017). Collectively, it has become apparent that neither CSCs are static entities, but rather affected by multiple factors throughout the lifetime of a cancer (Figure 3). Lastly, newly emerging views on the biology of normal stem cells suggest that CSCs do not necessarily have to be quiescent and rare, but they could also be abundant and vigorously proliferating, which will further challenge the simplistic hierarchical CSC model. (Kreso and Dick 2014; Cabrera, Hollingsworth, and Hurt 2015; Plaks, Kong, and Werb 2015; Batlle and Clevers 2017)

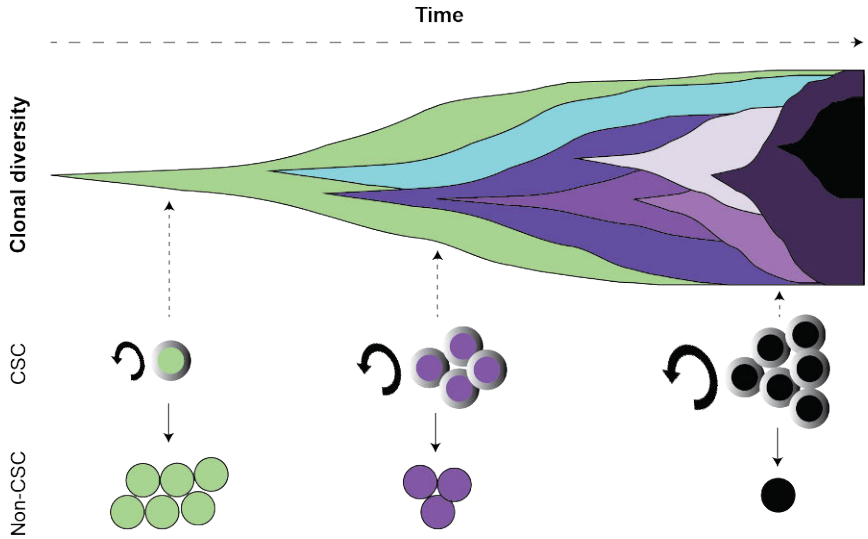

Figure 3. Unified model of cancer stem cells and clonal evolution, adapted from (Kreso and Dick 2014). Gain of mutations with selective advantage can result in clonal expansion of the founder cell (top panel, clonal diversity). In the beginning of cancer progression the hierarchy of a tumor is steep (left), with a minority CSCs existing among a large number of more differentiated non-CSCs. Over time, simultaneously, another cell can acquire a different advantageous mutation that enables it to form another subclone. Over the lifetime of a cancer, CSCs also evolve and acquire additional alterations that can influence CSC frequency. Some subclones (middle) may display an intermediate hierarchy, in which the number of CSCs is fairly high, but a hierarchy still exists. As tumors progress over time (right), mutations accumulate resulting in even higher capacity for self-renewal and further impairment of the remaining maturation programs. This then leads to further expansion of cells exhibiting CSC characteristics. In this model, tumor hierarchies become vague within the genetic subpopulations as cancers progress. Thus, the applicability of the CSC model in such homogeneous subpopulations is not anymore warranted.

Although considerable controversy remains regarding how to best define CSCs and which tumor types are genuinely hierarchically organized, the CSC model has gained reasonable interest as CSCs possess characteristics that make them clinically highly relevant. Even though CSCs have been shown to be a common property of various cancer subtypes and tumors from different tissues, their identification and eradication has not been as simple as was initially hoped. In many cancers, CSCs reside in dedicated niches as distinct populations that can be isolated from the bulk of the tumor cells and shown to have self-renewal and long-term clonal repopulation capacity. Evidence from both experimental models and clinical studies have demonstrated that CSCs can survive many commonly used cancer therapeutics and ionizing radiation. How CSCs are doing this, is not yet, however, clear. Nevertheless, the characteristics and transcriptional signatures specific to CSCs have proven to be highly predictive of overall patient survival in many hematological malignancies and

solid tumors that further highlight CSCs clinical relevance. Development of novel cancer therapies that are based on intrinsic CSC features, or modulation of the CSC niche functions, hold hope for improvement of survival of cancer patients. (Kreso and Dick 2014; Cabrera, Hollingsworth, and Hurt 2015; Plaks, Kong, and Werb 2015; Batlle and Clevers 2017)

2.2.1 CSCs in AML and CML

Leukemias have served as perfect models for studying the CSC hypothesis. The earliest *in vivo* evidence pointing to the existence of multiple functionally distinct cell subpopulations within single cancers came in the 1960s from AML and ALL patients (Gavosto et al. 1967; Clarkson et al. 1970). Further studies during the 1970s and 1980s provided observational evidence for the CSC hypothesis and denoted that not every leukemia cell was equal and only rare cells were leukemia inducing, later named as leukemia stem cells (LSC) (Clarkson et al. 1967; Fialkow 1974; Griffin and Löwenberg 1986). In order to state that a cancer is hierarchically organized, it was critical to demonstrate that cancer comprises of functionally discrete cell types that can be purified and assayed. The maturation of xenografting techniques in immune-deficient mice along with the development of fluorescence-activated cell sorting techniques, made engraftment of leukemic cells in mice feasible. These tools, coupled with quantitative assays, allowed the pioneering discoveries in the 1990s that provided the first evidence of LSCs in AML and CML patients (Lapidot et al. 1994; Bonnet and Dick 1997; Holyoake et al. 1999). Almost 30 years later, it is well acknowledged that both AML and CML arise from, and are maintained by, the LSCs.

2.2.1.1 Origins and definition of leukemia stem cells (LSCs)

There is increasing evidence in AML that the LCSs originate from the normal hematopoietic stem cells (HSCs) (Desai et al. 2018; Abelson et al. 2018). It has been shown that leukemogenesis relies on the sequential accumulation of somatic mutations within the multipotent and quiescent HSCs that then give rise to the pre-leukemic HSCs. The consequences of these "early" mutations in genes such as DNMT3A, IDH1/2 and TP53, are the enhancement or acquisition of self-renewal potential and impairment of differentiation (Corces et al. 2016), both of which can lead to variably expanding subclones of pre-leukemic HSCs (Shlush et al. 2014; Jan et al. 2012). These cells need further genetic and/or epigenetic events to develop into LSCs that then are capable of producing fully penetrant leukemic growth. "Late" mutations in proteins within signaling pathways such as FLT3, promote proliferation, enable a full differentiation block and drive the development of AML (Vetrie, Helgason, and Copland 2020). However, the situation in CP CML is highly

different as only a minority (15 to 20%) of patients with CP CML harbor pre-leukemic mutations (Kim et al. 2017) and as CP CML primarily arises from a single alteration, BCR-ABL1. That said, it has been demonstrated that also in CML, all patients have LSCs at diagnosis (Holyoake et al. 1999), but since the acquisition of BCR-ABL1 *per se* may impair self-renewal, and because most of the samples from patients with CML do not engraft well in immunocompromised mice, the definitions used in AML to define LSCs are not in most part compatible in CML. Overall, current evidence points to CP CML LSCs arising also from one or more specific subtypes of HSCs, but the mechanisms for this are not yet fully understood (Vetrie, Helgason, and Copland 2020). The view that there is actually not one but multiple subclones of LSCs and pre-leukemic HSCs present in one leukemia patient, is supported by the observation that both pre-leukemic HSCs and LSCs are characterized by extensively increased genetic instability (Gentry and Jackson 2013). Moreover, while normal HSC divide mainly in an asymmetrical manner (that is, producing one daughter cell), in LSCs the chance for symmetrical cell division that produce two identical cells is substantially increased (Gentry and Jackson 2013). Thus, the elevated chance of random genomic alterations in the pre-leukemic HSCs and LSCs genomes during the subsequent symmetrical cell divisions enables the generation of not only one but multiple subclones with distinct oncogenic drivers and immunophenotypes within a single patient.

It is important to explicitly distinguish LSCs from the healthy HSCs. For defining therapy naïve LSCs, cell surface marker analysis, biochemical assays to evaluate reactive oxygen species (ROS) levels, global molecular profiling and functional assays including engraftment into immunocompromised mice, and *ex vivo* co-culture systems have been utilized. The expression of CD34 and a lack of CD38 (CD34+CD38−) on the cell surface was the first immunophenotype associated with LSCs in both AML and CML (Lapidot et al. 1994; Holyoake et al. 1999), and the CD34+CD38− cells are still generally accepted as the main immunophenotype in which CP CML LSCs reside. Lately, however, the LSC phenotype has shown to be far more complex and numerous additional markers have been proposed for proper LSC immunophenotyping. These include CD9, CD25, CD33, CD47, CD93, CD99 and CD371 for both AML and CML, but the list seems to grow as the knowledge increases (Zagozdzon and Golab 2015; Vetrie, Helgason, and Copland 2020). Furthermore, engraftment has been considered as the gold-standard assay defining the LSCs. However, as there have been considerable improvements over the past decade in immunocompromised mouse strains due to removal of the requirement for irradiation of mice prior to the xenograft (McIntosh et al. 2015; Cosgun et al. 2014) and as there are strains that express human cytokines available (Wunderlich et al. 2010), increases in the proportion of AML samples that engraft and in the level to which they engraft have been detected (Vetrie, Helgason, and Copland 2020).

Extensive effort has been devoted to identifying genomic, transcriptomic and epigenetic signatures of LSCs within the past decade and at present the most explicit definitions of LSCs arguably lie within these global molecular profiles. Studies in AML have demonstrated that regardless of the immunophenotype or the oncogenic driver, LSCs exhibit distinct transcriptomic and epigenetic signatures that are founded on a small number of genes that are similar to, but discrete from, signatures identified in normal HSCs (Eppert et al. 2011; Jung et al. 2015; Ng et al. 2016). These signatures have very good prognostic value and they have already served as a basis for drug discovery for identification of compounds that target AML LSCs (Laverdière et al. 2018). Transcriptomic and epigenetic profiling has also been carried out on LSCs from patients with CML (Abraham et al. 2016; Scott et al. 2016), but the limited scope of these studies with small sample cohorts while bearing in mind the issues associated with engrafting CML samples in mice calls for further research on this topic before any predictors of the clinical outcome in CML can be declared.

2.2.1.2 LSCs and disease relapse

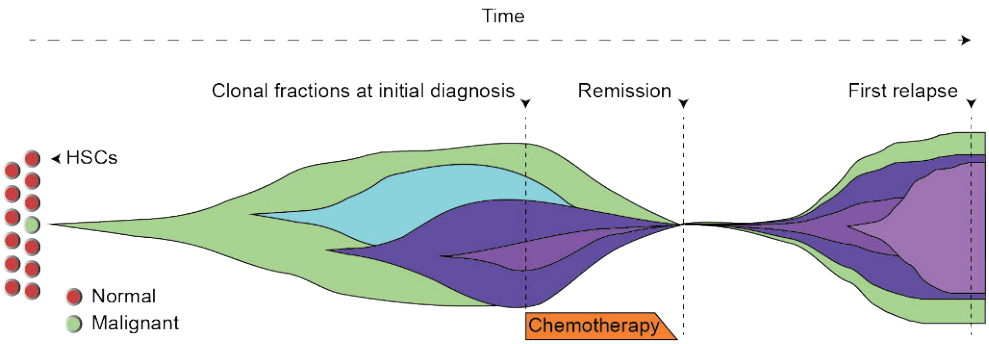

Figure 4. Graphical representation of clonal evolution from the primary tumor to relapse in AML.

LSCs have been shown to persist in bone marrow as biologically distinct subpopulations during remission of AML and CML (Vetrie, Helgason, and Copland 2020). However, multiple LSC subclones detected in most of the patients at AML relapse, point towards continued evolution in LSCs even during remission. At relapse (Figure 4), founder clones already present at diagnosis can re-emerge through branching evolution with novel sub-clonal structures, or clones that were minor populations at diagnosis may arise as dominant ones (Ding et al. 2012; Shlush et al. 2017). In the majority of cases, disease relapse has been shown to arise from the latter ones. However, thus far pre-leukemic HSCs have not been shown to be the origin of the leukemic clones driving relapse. In multiple patients, novel mutations

have been reported to be present in the relapse clones, suggesting that the exposure to chemotherapy itself induces mutagenesis that further accelerates LSC clonal evolution (Ding et al. 2012; Shlush et al. 2017). The situation, however, is again different in CML as TKIs *per se* do not cause DNA damage that would directly drive LSC clonal diversification. Moreover, even though most patients with CP CML on TKI treatment remain in stable remission, also in CML disease relapse and progression occur and typically the underlying reason is the emergence of Ph chromosome positive clones carrying mutations in BCR-ABL1 kinase domain or other genes linked to TKI resistance (Soverini et al. 2015; Kim et al. 2017). It has been postulated that the appearance of these mutations during TKI treatment is most likely linked to TKI-induced selective pressure that enriches for TKI-resistant LSC clones that carry mutations which arose already early in leukemogenesis or even as pre-leukemic alterations (Kim et al. 2017; Branford et al. 2018). This would then even propose that the outlook for many CML patients might be predetermined even before the first-line treatment begins, which again holds not true for most patients with AML.

2.2.2 CSCs in solid tumors

CSCs in solid tumor were for the first time identified in breast cancer (Al-Hajj et al. 2003). This achievement illustrated that the same principles regarding CSCs that had previously been demonstrated to apply in leukemia, could also be translated to solid tumors. After the discovery in breast cancer, CSCs have been identified in various solid cancers including prostate, lung, HNSCC, colorectal, and brain, just to name a few (Batlle and Clevers 2017). Recent results in solid tumors imply that transcriptional, metabolic and epigenetic signatures specific to CSCs are highly prognostic for disease outcome across a wide variety of patients with varying driver mutations (Merlos-Suárez et al. 2011; Smith et al. 2018; Pece et al. 2019; Intlekofer and Finley 2019). At the moment, combinatorial approaches that integrate publicly available data from numerous sources are increasingly conducted to produce CSC and stemness signatures or indexes, that aim to improve patient stratification of patients with cancer irrespective of the primary site of the disease.

Immunophenotyping of CSCs in solid cancers has shown to be highly context and cancer type dependent. Phenotypic heterogeneity within CSC subpopulations exists and the CSC phenotype is not equal between cancer subtypes or even tumors of the same subtype. Cell surface markers most frequently used for CSC isolation in solid tumors include: CD44, CD24, nestin, CD34, CD133, and EpCAM (Visvader and Lindeman 2008; Shimokawa et al. 2017). However, as CSCs from solid tumors have frequently been extracted using markers specific for normal stem cells of the same organ, none of these markers gain exclusive extraction of CSCs. Partly due to

this, the fraction of CSCs in solid tumors and the pure hierarchical organization of these tumors remains debatable. Moreover, a great variety exist in the frequencies of CSCs reported in solid tumors reflecting both the genuine biological variation as well as technical issues. These technical issues include the purity of tumor cell fractionation, the manipulation needed to dissociate solid tissues into single cells, and the challenges associated with xenotransplantation including the lack of an appropriate microenvironment in the immunocompromised mice as well as the lack of an intact immune system. For example, in HNSCC it has been estimated that CSCs account for only 1 to 5% of the HNSCC cell population (Hutchinson, Mierzwa, and D'Silva 2020), whereas on the contrary in melanoma, depending on the technology applied, the frequency varies from 1 % to 25% (Quintana et al. 2008). Furthermore, there is growing evidence supporting the concept of CSC plasticity (Cabrera, Hollingsworth, and Hurt 2015; Gupta et al. 2019) in which the microenvironment plays a central role in shaping the CSC phenotype towards temporal and spatial heterogeneity. For example, in glioblastoma, it has been reported that cells displaying glioma CSC-associated membrane markers do not represent a clonal entity that is defined by distinct transcriptomic profiles and functional properties, but rather a plastic cellular state that most cancer cells are able to adopt (Dirkse et al. 2019). Thus, while the concept of CSCs as a rare cancer cell population that exhibit stem cell like characteristics and promote the growth of hematological malignancies is widely accepted, the hierarchical CSC model has been challenged in solid cancers and it seems that the situation is far more complex in many solid malignancies.

As previously discussed, the definition of CSC in solid cancers is blurry and matters of controversy remain. Maybe due to this, the terms “stemness”, “cancer stem cell like cells” and “cancer progenitor cells” are increasingly being used in the literature to refer collectively to the cellular phenotype and functions that control and maintain the stem cell state in cancer. At the moment, it is, however, unclear whether the stemness phenotype observed in many cancers reflects the presence of real CSCs or simply the hijacking of stem cell-associated programs by the non-CSC cancer cells, or both. Whatever the fundamental mechanism may be, stemness has been denoted as an important phenomenon due to its strong association with poor outcome in a wide spectrum of cancers (Ng et al. 2016; Tirosh et al. 2016; Smith et al. 2018; Malta et al. 2018). Furthermore, similarities between CSCs and embryonic stem cells (ES) have been observed and transcription factors highly expressed in ESs, such as OCT4, SOX2 and NANOG, have been reported to be re-expressed or reactivated in CSCs. For example, OCT4, Octamer-binding transcription factor 4, has been used to characterize CSCs in multiple solid cancers. OCT4 is a *bona fide* stem cell marker and thus a cardinal regulator of self-renewal, pluripotency and maintenance of the stem cells. Interestingly, several studies have shown that OCT4 is highly expressed also in CSCs of other tumors than those of embryonal origin, such as lung (CD133+

subpopulation, (Chen et al. 2008)) and breast cancer (CD44+/CD24− subpopulation, (Ponti et al. 2005)). Furthermore, elevated OCT4 expression in cancer has been linked to CSC phenotype (Ponti et al. 2005), radiation resistance (Murakami et al. 2015) and poor prognosis (Chiou et al. 2008; Shen et al. 2014). However, it might be that the elevated expression of OCT4 and the other "stemness genes" is rather a function of malignant dedifferentiation and not thus unique to the CSC population. Nevertheless, regardless of the underlying mechanism, machine learning and omics data analysis by integrating information from publicly available transcriptomic and epigenomic data have been recently utilized to create cancer stemness indexes (Malta et al. 2018; Miranda et al. 2019). These stemness indexes do not discriminate between CSCs and the bulk of the tumor cells, but with this approach, indexes that can be beneficial in defining metastatic tumors and revealing potential drug targets for anti-cancer therapies have been identified. These findings further underline the central roles of stemness and oncogenic dedifferentiation in cancer.

2.3 Protein phosphatase 2A (PP2A)

Protein phosphorylation is a reversible post-translational modification, that is conducted by the counteracting phosphate group adding protein kinases and the phosphate group removing protein phosphatases. Protein phosphorylation is a rapid and efficient mechanism for modulating protein function and the most common post-translational modification of proteins (Khoury, Baliban, and Floudas 2011). In the majority of the cases, a phosphate group is covalently bound or removed from serine, threonine or tyrosine amino acid residues and to lesser extent from histidine, lysine or arginine residues of a substrate protein. Phosphorylation alters the charge and conformation, both local and global, of the substrate protein resulting in modulation of the stability, function and subcellular localization of the protein. Multiple cellular signaling cascades are finetuned by the phosphorylation events as alterations in phosphorylation status enables cells to adapt to changing circumstances quickly. Phosphorylation-dependent cellular signal transduction is highly controlled process that requires coordinated and timely regulation of both kinases and phosphatases and the deregulation in this balance underlies many human diseases, including cancer (Hunter 1995).

The human genome encodes more than 500 protein kinases and less than 200 protein phosphatases (Manning et al. 2002; Fleuren et al. 2016; Duan, Li, and Köhn 2015; Chen, Dixon, and Manning 2017). Protein kinases and phosphatases have been denoted to specific subcategories based on their substrate specificity. About 400 protein kinases carry out phosphorylation at serine or threonine residues, thus classified as serine/threonine kinases, whereas the remaining protein kinases are mostly classified as tyrosine kinases (Fleuren et al. 2016; Shi 2009). A very small fraction of protein kinases preferably phosphorylate histidine, lysine and arginine

residues and are thus referred to as histidine, lysine and arginine kinases (Matthews 1995). The same substrate specificity-based classification is applied for protein phosphatases, although the evolutionary origins of phosphatases are more diverse than of kinases, leading to a structurally far more heterogeneous group of enzymes (Li et al. 2013; Chen, Dixon, and Manning 2017). Nevertheless, the approximately 200 known protein phosphatases can be divided into four main classes: 1) serine/threonine phosphatases, 2) tyrosine phosphatases, 3) dual specificity phosphatases, and 4) histidine phosphatases (Shi 2009; Chen, Dixon, and Manning 2017). Although the number of protein phosphatases is smaller than of opposing kinases, the structure of phosphatase complexes, in which several mutually exclusive regulatory subunits dictate the substrate specificity and subcellular localization, enables a single catalytic subunit of a phosphatase to be part of even hundreds of phosphatase holoenzyme complexes. Hence, the total number of protein phosphatase complexes undoubtedly outnumbers the protein kinases in a cell.

2.3.1 Structure and function of PP2A

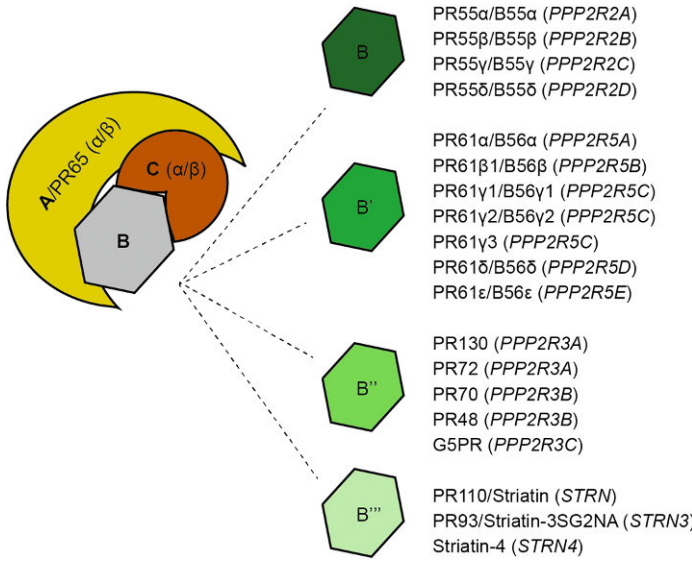

Figure 5. Schematic illustration of the PP2A holoenzyme structure. A is the scaffolding subunits, C is the catalytic subunit and the four regulatory subunit families, B, B', B'' and B''', with their nomenclature are presented. Encoding genes in brackets.

One of the major constituents of the total cellular serine/threonine phosphatase activity is the highly conserved and ubiquitously expressed protein phosphatase 2A, PP2A. Heterotrimeric PP2A holoenzyme consists of three distinct subunits: a

scaffolding A-subunit, a regulatory B-subunit, and a catalytical C-subunit (Figure 5). The main function of A-subunit is to act as a scaffold for the binding of C- and B-subunits and thus to enable assembly of PP2A complexes in spatial-temporal manner. The C-subunit contains the active site needed for the catalytic phosphatase activity, in which the hydrolysis of the serine/threonine phosphate esters in the presence of two manganese ions is conducted. The B-subunits contain substrate binding sites, and thus function as regulatory subunits whose binding directs the holoenzyme activity towards a specific set of substrates. All of the subunits have multiple isoforms, and the expression of any given isoform is cell and tissue dependent. The scaffolding and catalytic subunits each have two isoforms, α and β, which share high sequence similarity but are encoded by two distinct and functionally non-redundant genes (*PPP2R1A, PPP2R1B* for the A and *PPP2CA, PPP2CB* for the C). Out of these, Aα and Cα are predominant in the majority of the cell types in adult tissues, whereas Aβ isoform expression is elevated during the early stages of vertebrate development. However, functional complexity of PP2A emerge mainly via the regulatory B-subunits, which determine both the substrate specificity and subcellular localization of the heterotrimers. PP2A B-subunits are encoded by four families of genes (Figure 5, B, B', B'' and B'''), comprising a total of fifteen genes that further generate at least 26 different B-subunits, when all the isoforms and the known splice variants are counted. The sequence similarity among the four families is very low, and the expression levels of different B-subunits are highly diverse depending upon cell, tissue and developmental context. Thus, PP2A is not a single entity but a family of heterotrimeric holoenzymes with context dependent functions. Based on the number of known scaffolding, catalytical and regulatory subunits, PP2A can theoretically exist as 96 distinct holoenzyme compositions in human cells. (Janssens and Goris 2001; Xu et al. 2006; Eichhorn, Creyghton, and Bernards 2009; Sangodkar et al. 2016)

PP2A is required for the appropriate function of numerous signaling pathways involved in development, cell cycle regulation, cell growth control, DNA damage response, stress response, cell adhesion, cell mobility, and apoptosis (Janssens and Goris 2001; Xu et al. 2006; Eichhorn, Creyghton, and Bernards 2009; Sents et al. 2013; Sangodkar et al. 2016; Fowle, Zhao, and Graña 2019). As PP2A influences virtually all aspects of cell biology, its function must be strictly regulated in order to maintain cellular homeostasis. To prevent the formation of catalytically active complexes that lack correct substrate specificity, the assembly of holoenzymes *per se* forms a level of tight regulation (Sents et al. 2013). Unpartnered C-subunits are directed to ubiquitination and proteasomal degradation. Furthermore, PP2A stability and activity is regulated by multiple post-translational modifications and interacting proteins. For example, the phosphorylation of threonine 304 and tyrosine 307 of the C-subunit has been shown to inactivate PP2A and also to have a role in the selection of B-subunit

binding (Chen, Martin, and Brautigan 1992). Additionally, reversible methylation of leucine 309 of the C-subunit is linked to modulation of holoenzyme assembly as well as phosphatase activity of the PP2A complex (Ogris et al. 1999; Longin et al. 2007; Nunbhakdi-Craig et al. 2007; Stanevich et al. 2011). Methylation of leucine 309 is needed for B55-family subunit binding and enhances binding of some B56-family subunits, and it is also required for the active conformation of the C-subunit. Other post-translational modifications have also been reported, but the effects of these on PP2A function need more clarification. The deregulation and dysfunction of PP2A in disorders such as cancer, neurodegenerative disorders and diabetes have broadened our understanding of PP2A function both in health and disease.

2.3.2 PP2A in cancer

The first indication for PP2A's tumor suppressor functions originate from studies with okadaic acid, a carcinogenic shellfish toxin, that was demonstrated to selectively inhibit the phosphatase activity of PP2A and to induce tumor growth (Suganuma et al. 1988; Bialojan and Takai 1988). The tumor suppressor role of PP2A was further confirmed by findings implying that PP2A is one of the main targets for various DNA tumor virus proteins, such as adenovirus E4orf4, polyomavirus small and middle T antigens, and SV40 ST (Pallas et al. 1990; Shtrichman et al. 1999). Collectively these viral proteins function by displacing B-subunits from the PP2A holoenzyme, which then leads to altered PP2A activity towards distinct substrates (Eichhorn, Creyghton, and Bernards 2009; Westermarck and Hahn 2008). As mentioned earlier, importantly, studies with SV40 ST also identified inhibition of PP2A as one of the prerequisites for the cellular transformation of human cells into cancer cells. However, not all PP2A complexes are considered tumor suppressive. In cellular transformation, the importance of inactivation of PP2A has been associated with its ability to function as an antagonist of certain oncogenic signaling pathways. In fact, it was demonstrated that among the various PP2A heterotrimer complexes, only specific inhibition of complexes containing B56α, B56γ and PR72/PR130 subunits contribute to the PP2A-inhibition mediated tumor suppression and the perturbation of these complexes was shown to result in activation of MYC, WNT and PI3K/AKT pathways (Sablina et al. 2010). That said, as PP2A stands in the crossroad of multiple oncogenic signaling cascades, it is obvious that the tumor suppressor activities of PP2A also act through additional signaling pathways than the ones mentioned above. Indeed, very recent phosphoproteome analyses have revealed a large number of novel cancer relevant targets for the PP2A/B55 and PP2A/B56 tumor suppressive complexes (Kauko et al. 2020; Kruse et al. 2020). However, based on the accumulating cancer genomics data, it has been even postulated that the prime tumor suppressor function of PP2A might

actually be linked to the maintenance of genome stability instead of the regulation of oncogenic pathways implicated by the experimental transformation models (Kauko and Westermarck 2018).

Due to the emergence of the large-scale cancer genomics data within the past decade, the landscape of cancer related genomic alterations of PP2A have been brought into daylight only rather recently. Interestingly, current evidence implies that PP2A inhibition in cancer by non-genetic mechanisms such as altered expression of PP2A inhibitor proteins and post-translational modifications of PP2A subunits, is far more common than dysregulation of PP2A caused by genomic aberrations (Kauko and Westermarck 2018). Deregulation of PP2A subunit genes has been identified in some cancers, but the only two recurrently reported mechanisms include aberrant splicing of *PPP2R1B* (Aβ) and inactivating methylation of *PPP2R2B* (B55β) promoter (Meeusen and Janssens 2018). It seems that the cancer associated genetic and epigenetic alterations are more likely to affect genes encoding A- and certain B-subunits than the catalytic subunits of PP2A. Thus, in the majority of the cases, they only impair a subset of PP2A's functions. Furthermore, most copy number alterations of the PP2A subunit genes can be simply linked to their chromosomal location as they mainly occur as co-amplifications or deletion with other genes from the same location. However, two recurrent alterations with sturdy proof from comprehensive cancer genomics studies have been identified, the homozygous deletions of *PPP2R2A* (B55α) gene and point mutations in *PPP2R1A* (Aα) gene (Vogelstein et al. 2013; Lawrence et al. 2014; Zack et al. 2013; Kauko and Westermarck 2018). In some cancers, *PPP2CB* (Cβ) is also frequently co-deleted with *PPP2R2A* gene, which can be explained by their close chromosomal location. At the moment, however, it is unclear whether the deletion of *PPP2CB* offers any significant selective advantage to cancer cells as it is known that the Cα subunit is about 10-fold more abundant in cells than Cβ (Fowle, Zhao, and Graña 2019), and as the depletion of Cβ does not substantially contribute to malignant transformation whereas even partial depletion of the Cα fosters the transformation (Sablina et al. 2010). Observations of the above-mentioned recurrent and significant aberrations have been reported also in hematological malignancies, excluding mutations in *PPP2R1A* gene that seem to be an uncommon mechanism in AML (TCGA and Beat AML datasets), ALL (St Jude dataset) and CLL (Broad Institute dataset) (no data available for CML at the moment) based on the data and statistical calculations retrieved from cBioportal (Cerami et al. 2012). Low frequency homozygous deletions in B55α have been documented in primary plasma cell leukemia (Mosca et al. 2013), and nonsense mutations in B55α have been reported to contribute to severe protein truncations and loss of B55α protein expression in AML (Shouse et al. 2016). Furthermore, B-cell chronic lymphocytic leukemia (B-CLL) is one of the diseases affected by the reduced *PPP2R1B* gene expression

caused by aberrant splicing of the mRNA (Kalla et al. 2007). Of a note, no significant deregulation of any of the PP2A B55- or B56-family subunits was detected in the Beat AML program cohort (Tyner et al. 2018) including 672 tumor specimens collected from 562 patients, when expression data and statistical calculations was retrieved from cBioportal.

2.3.2.1 PP2A inhibitor proteins (PAIPs) involved in cancer

The non-genomic inhibition of PP2A activity by elevated expression of endogenous PP2A inhibitory proteins (PAIPs) greatly exceeds in magnitude and effect the low frequency genetic alterations observed in PP2A genes in human cancers. Many of these endogenous inhibitors function in certain phases of the cell cycle and regulate only a subset of PP2A targets. Several functionally and structurally diverse proteins have been designated as PAIPs based on their ability to interact with PP2A, and to inhibit the phosphatase activity of PP2A. Many of the PAIPs have been demonstrated to also have other, PP2A independent functions. As several of the PAIPs have been found overexpressed in various human cancers, their inhibition could achieve indirect PP2A reactivation and hence provide a potent anti-cancer approach. The PAIPs shortly discussed here are: Cancerous inhibitor of PP2A (CIP2A), cAMP-regulated phosphoprotein 19 (ARPP19), α-endosulfin (ENSA), SET (I2PP2A), Protein phosphatase methyl esterase 1 (PME-1) and Type 2A interacting protein (TIPRL) (Figure 6).

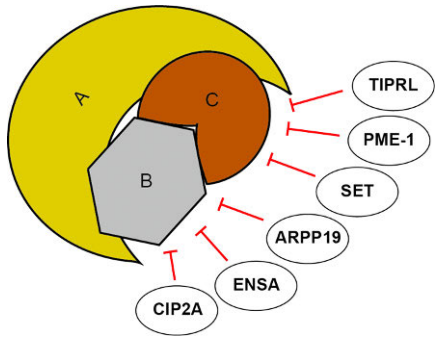

Figure 6. Schematic representation of PP2A inhibitory proteins discussed in this chapter.

CIP2A

CIP2A functions as an oncoprotein by directly binding to PP2A subunits B56α and B56γ as an obligatory homodimer (Wang et al. 2017; Soofiyani, Hejazi, and Baradaran 2017). CIP2A is expressed at low levels in non-proliferating normal cells, except in testis, and overexpression has been observed virtually in all cancers studied

(Junttila et al. 2007; Ventelä, Côme, et al. 2012; Khanna and Pimanda 2016). The frequency of CIP2A overexpression is exceptionally high (40 to 90% of patient specimens) in many human cancer types and high CIP2A protein expression correlates with cancer progression and poor patient survival in a wide spectrum of human malignancies including AML (Barragán et al. 2015), CML (Lucas et al. 2011) and HNSCC (Junttila et al. 2007). Overexpression of CIP2A has also been reported to broadly promote cancer cell drug resistance (Laine et al. 2013; Kauko et al. 2018). In many cancers, CIP2A mediated PP2A inhibition fosters stabilization of phosphorylated MYC at serine 62, as well as activation of AKT signaling pathway (Junttila et al. 2007; Chen et al. 2010). Hence, CIP2A has been shown to contribute to malignant cellular growth and tumor formation *in vitro* and *in vivo*. Additional modes of action for CIP2A in promoting aggressive cellular growth include inhibition of apoptosis via inhibited dephosphorylation of PP2A substrate death associated protein kinase (DAPk) (Guenebeaud et al. 2010) and prevention of PP2A mediated growth arrest and senescence via positive feedback loop with E2F1 (Laine et al. 2013). In addition to E2F1 and MYC, CIP2A expression in cancer cells has shown to be regulated at least by DNA damage checkpoint kinase (CHK1) and EGFR/MEK/ETS1 pathway (Khanna et al. 2013; Khanna, Pimanda, and Westermarck 2013). Importantly, as CIP2A depletion does not compromise normal mouse growth or development (Ventelä, Côme, et al. 2012; Laine et al. 2013), inhibition of CIP2A protein expression or protein activity could be utilized in the future in development of highly specific cancer therapy with only minor side effects.

ARPP19 and ENSA

ARPP19 and ENSA are ubiquitously expressed PP2A inhibitors that have been demonstrated to promote G_2/M transition and the mitotic state by inhibiting PP2A activity towards CDK1-phosphorylated substrates (Gharbi-Ayachi et al. 2010; Mochida et al. 2010). ARPP19 and ENSA are members of an evolutionary conserved protein family but are generated from distinct genes and mainly differ within their 20 amino acid N-terminal domain. ARPP16, a splice variant of ARPP19, has also been identified and shown to be highly enriched only in specific brain areas (Girault et al. 1990). Rather recently ARPP16 was identified as a PP2A inhibitor in mammalian brain when phosphorylated at serine 46 by MAST3 kinase (Andrade et al. 2017; Musante et al. 2017). At present, most of the literature related to ARPP19 and ENSA mediated PP2A inhibition have been concentrating on their role in cell cycle and especially in mitosis. Both ARPP19 and ENSA have shown to have mutual and separate, and even oncogenic functions. In mammalian cells, B55α and B55δ containing PP2A complexes are the main targets of these inhibitor proteins (Manchado et al. 2010). Both B55α and B55δ are ubiquitously expressed in the body.

Prior to the binding of ARPP19 and ENSA to PP2A, a phosphorylation of serine 62 on ARPP19 and serine 67 on ENSA is required, and this is conducted by the mitotic Greatwall kinase (GWL or MASTL) (Gharbi-Ayachi et al. 2010; Mochida et al. 2010). However, the structural details of how ARPP16, ARPP19 and ENSA interact with and inhibit PP2A are not yet known. Interestingly, whereas ARPP19 knock-out is embryonically lethal in mice, depletion of ENSA is not, suggesting for their differential roles in development (Hached et al. 2019). Furthermore, ENSA's role as oncoprotein remains debatable as in some context ENSA has been associated with tumor suppression and in other with oncogenic functions (Chen et al. 2013). However, consistent with its functional role as an inhibitor of PP2A, ARPP19 overexpression has been linked to the cancer progression and it correlates with tumor grade both in human glioma (Jiang et al. 2016) and hepatocellular carcinoma (Song et al. 2014). Moreover, cell proliferation was reported to be increased following ARPP19 overexpression in normal breast cells (MCF10A cells) (Vera et al. 2015). Also, in gastric cancer, high ARPP19 expression positively associated with Herceptin resistance and poor patient survival rate (Gao et al. 2020), whereas in breast cancer cells, ARPP19 was reported to be a target of microRNA-320a and mediate tamoxifen resistance (Lü et al. 2015). Interestingly, functions of ARPP19 have been reported to be regulated by several different miRNAs in multiple cancer types (Bruchova-Votavova, Yoon, and Prchal 2010; Lü et al. 2015; Ye et al. 2020; Ma et al. 2020; Gong et al. 2018). Neither the oncogenic role of ARPP19 nor ENSA has yet been studied in hematological malignancies.

SET

SET inhibits PP2A activity by direct binding to the catalytic C-subunit of PP2A (Al-Murrani, Woodgett, and Damuni 1999). Recently SET was also shown to associate with B56α in gastric cancer (Enjoji et al. 2018), suggesting specific inhibition of the PP2A-B56 complex in this cancer. Two highly similar transcript variants of SET, SETα and SETβ, have been identified and both of the resulting protein isoforms exhibit PP2A inhibitory activity (Saito et al. 1999). SET is a phosphoprotein and phosphorylation of serine 9 of SET has shown to be mandatory for PP2A inhibition (ten Klooster et al. 2007). SET expression is modulated as a function of cell proliferation, with low expression in quiescent or contact-inhibited cells, and high expression in rapidly dividing cells or transformed cells (Shin et al. 1999). Even though SET is undoubtedly a valid inhibitor of PP2A, several functions have been described for SET and many of these are conducted via PP2A-independent mechanisms, as for example the negative regulation of tumor suppressor p53 (Kim et al. 2012). To date, SET has been implicated in various cellular processes such as differentiation, transcription, DNA replication, chromatin remodeling, cell cycle

regulation and cell migration (Li, Makkinje, and Damuni 1996; Seo et al. 2001; Kandilci, Mientjes, and Grosveld 2004; ten Klooster et al. 2007). Interestingly, high SET expression resulting in progressive loss of PP2A activity has been reported in several hematological malignancies with therapeutic implications. SET expression correlates with disease progression and high SET is associated with poor prognosis in AML (Cristóbal et al. 2012), BCR/ABL-positive CML (Neviani et al. 2005), BCR/ABL-positive ALL (Neviani et al. 2007), and B-CLL (Christensen et al. 2011). In BCR/ABL positive cell lines and in patient-derived cells, SET expression has been shown to be enhanced by BCR/ABL (Neviani et al. 2005). Data from multiple preclinical studies imply that PP2A reactivating drugs, including OP449 and FTY720, possess therapeutic potential in CML and other BCR/ABL positive leukemias (Neviani et al. 2007; Arriazu, Pippa, and Odero 2016; Agarwal et al. 2014). These drugs have been shown to inhibit the SET-PP2A interaction and consequently to impair BCR/ABL leukemogenic potential by causing inactivation and downregulation of the BCR/ABL oncoprotein itself. A fusion protein of SET with nucleoporin Nup214 (also known as CAN) resulting from recurrent chromosomal rearrangement or translocation observed in AML and T-ALL, has been additionally shown to support leukemia progression (Van Vlierberghe et al. 2008; von Lindern et al. 1992).

PME-1

PME-1 has multiple modes of action for inhibiting PP2A activity. PME-1 demethylates leucine 309 on PP2A catalytic C-subunit and is opposed by the methylating enzyme LCMT1 in mammalian cells (Lee and Stock 1993; Ogris et al. 1999; Kaur and Westermarck 2016). As mentioned previously, this methylation plays also a role in regulating PP2A holoenzyme assembly and whereas methylation increases the activity of PP2A, in contrast, demethylation and binding of PME-1 to the active site of PP2A results in inactivation of PP2A activity. In addition, PME-1 functions as a stabilizer of a native inactive PP2A dimer until it is activated and eventually assembled into a functional holoenzyme (Longin et al. 2004). Furthermore, PME-1 has been shown to remove manganese ions required for the catalytic activity of PP2A while bound to the active site (Xing et al. 2008). In the context of cancer, elevated PME-1 expression has been reported in endometrial cancer and glioblastoma, and in both, it correlates with cancer stage (Wandzioch et al. 2014; Puustinen et al. 2009). The oncogenic role of PME-1 has been well studied in human glioma, where PME-1 has been shown to regulate the MAPK/ERK signaling pathway via PP2A (Puustinen et al. 2009). Moreover, in a recent study, PME-1 expression was shown to confer therapy resistance in gliomas and reactivation of PP2A by PME-1 depletion was shown to sensitize glioma cells to

multikinase inhibitors (Kaur et al. 2016). The role of PME-1 in hematological cancers remains to be investigated.

TIPRL

TIPRL (or TIP or TOR signaling pathway regulator) binds to the catalytic C-subunit of PP2A and mutations in Aα-subunit have been shown to enhance this interaction (McConnell et al. 2007; Haesen et al. 2016). In endometrial cancers, TIPRL has been shown to even preferentially bind to the mutant Aα frequently found in these cancers (Haesen et al. 2016; Wu et al. 2017). Interestingly, TIPRL was recently shown to tolerate disease-associated PP2A mutations in A-subunit during their interaction, unlike the overlapping B-subunits, resulting in reduced holoenzyme assembly and enhanced inactivation of the mutant PP2A (Wu et al. 2017). The oncogenic role of TIPRL is thus only emerging and what is at the moment known about TIPRL's cellular functions is that it has an important role in mammalian TOR signaling as well as DNA damage response (Nakashima et al. 2013; McConnell et al. 2007). Upregulated expression of TIPRL has been documented in hepatocellular cancer, where it was shown to protect cancer cells from apoptosis (Song et al. 2012). The role of TIPRL in other cancers remains unknown.

2.4 Cancer biomarkers

According to the National Cancer Institute (NCI), biomarker is defined as: "a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease" (www.cancer.gov). At the moment though, organizations, institutes and publications use different definitions for a biomarker. In medicine, the term is used to cover any physiological, biochemical, molecular, or anatomical quality that can be measured or quantified (Califf 2018). A cancer biomarker integrates all the attributes mentioned above and refers usually to a process or substance that is indicative of the presence of cancer in the body. In most of the cases this means a specific alteration at the DNA, RNA or protein level that can then be used as a genetic, epigenetic, proteomic, glycomic or imaging biomarker. Ideally, a cancer biomarker should be such that it could be assayed in non-invasively collected sample materials such as blood or urine. On the other hand, only after a cancer biomarker is validated, can it serve in clinical practice. A huge "valley of death" remains between the promising biomarkers that arise from biomedical research and the validated biomarkers that can reliably serve in clinical practice and help clinicians in decision making.

Biomarkers are used in four primary ways in cancer medicine and research: as diagnostic, prognostic, monitoring and predictive (Califf 2018). A single biomarker

may serve in one or all of these phases, simply depending on its nature. A diagnostic cancer biomarker identifies or confirms the presence of a cancer of interest. Such biomarkers can be used for risk assessment when pinpointing individuals that are predisposed to particular types of cancers or for screening for the presence of a cancer at an early stage. Diagnostic biomarkers can also be used in narrowing down diagnosis to a specific subpopulation of patients. A prognostic cancer biomarker, on the other hand, can serve in forecasting the aggressiveness of an identified cancer as it is used to identify the likelihood of disease progression or disease recurrence. A monitoring biomarker is a biomarker that can be measured repeatedly to assess the status of a disease, or to monitor the efficacy of a given therapy (Califf 2018). Lastly, a predictive cancer biomarker can help in deciding the optimal treatments for an individual as it discriminates those who will and will not respond to any given therapy. (Robb, McInnes, and Califf 2016; Califf 2018)

In the era of high-throughput omics technology, thousands of biomarker candidates can be investigated in-parallel easily. In terms of PP2A biology, multiple promising cancer biomarkers have been identified by basic research. For example, CIP2A would have potential to be utilized as a diagnostic, prognostic and predictive biomarker in several cancers. As levels of CIP2A are very low in normal cells and overexpression has been observed virtually in all cancers studied (Junttila et al. 2007; Ventelä, Côme, et al. 2012; Khanna and Pimanda 2016), CIP2A could serve as a potential diagnostic biomarker. On the other hand, high CIP2A expression correlates with cancer progression and poor patient survival in a broad spectrum of human cancers (Khanna and Pimanda 2016) and thus CIP2A could also be used as a prognostic biomarker. Overexpression of CIP2A has also been reported to broadly promote cancer cell drug resistance (Laine et al. 2013; Kauko et al. 2018), and thus CIP2A could be a strong candidate to be used as a predictive biomarker. However, only after a biomarker is properly validated, can it serve in prospective research or clinical practice.

A number of challenges exist in the biomarker development process, and during the bench-to-bedside development that covers four distinct phases of biomarker development; discovery phase, confirmation phase, validation and refinement phase, and adoption phase, a series of strict processes must be undertaken (Sawyers and van 't Veer 2014; Parkinson et al. 2014). When evaluating biomarkers, it would be crucial to understand that an association or correlation between a measured biomarker level and a clinical event does not directly mean that the biomarker can be considered as a valid surrogate. However, this is one of the most common and serious errors made in the biomarker evaluation (Robb, McInnes, and Califf 2016). For a biomarker to qualify as a surrogate, the change in the biomarker should be able to explain the change in the clinical event. The term “explain” involves statistical analysis that can only be conducted with confidence if the main finding can be reproduced in several

independent studies and if the basic biology behind the biomarker is well understood. Due to this, most of the promising biomarkers are not valid surrogates and even if a biomarker can be validated as a surrogate, it has to be kept in mind that this validation holds true only in a distinct context of use (Robb, McInnes, and Califf 2016).

The amount of work needed to validate a biomarker is substantial. Many initiatives worldwide have been taken to harmonize the terms used in translational science and medical product development, such as the BEST resource (Biomarkers, EndpointS, and other Tools, https://www.ncbi.nlm.nih.gov/books/NBK326791/) by the joint leadership conference of the U.S. Food and Drug Administration (FDA) and the National Institutes of Health (NIH). Also, initiatives to compile a list of criteria by which biomarkers should be assessed in order to streamline clinical validation, such as the EDRN initiative (Early Detection Research Network, https://edrn.nci.nih.gov) by the National Cancer Institute (NCI) have been taken. Both BEST and EDRN guidelines are constantly being updated and both of them aim to accelerate the translation of promising biomarker discoveries into approved medical products and applications. Validation criteria listed by the EDRN for biomarkers include proof of concept, experimental validation, analytical performances validation and protocol standardization. Throughout the pipeline of biomarker research and application development, commitment to stringent transparency and reproducibility is essential. One of the most important steps in the biomarker validation is the assay's analytical performances validation that should include at least sensitivity, specificity, robustness, accuracy, reproducibility, practicality and ethicality aspects. The main goal should be to determine an assay that guarantees that the biomarker can be measured repeatably, precisely and reliably in a timely manner and at a low cost. If assays are not validated properly, misbelief about the biomarker's value can be created and also the ability to match a biomarker with its' appropriate usage can get hindered. Also, when resources with mistaken concepts about future use are allocated to the development of a biomarker application that does not meet the criteria for clinical use or even for regulatory approval, the money and time invested are being wasted. Although many of the issues listed above may be considered self-evident, they are too often neglected. Hence, before a biomarker can reach the clinical practice, there are multiple considerations that need to be taken, and the high number of challenges around biomarker development means that only very few markers, according to some estimations even as low as 0.1%, in the end achieve a substantial clinical role (Goossens et al. 2015).

3 Aims of the Study

In normal cells PP2A functions as a growth suppressor, but in cancer cells its functions are mainly inhibited by the endogenous inhibitor proteins CIP2A, PME-1, SET, ARPP19 and TIPRL. Expression of these PAIPs have been studied in multiple cancer types, but their status in hematological cancers is still unclear and their co-expression in myeloid leukemias is not known. Furthermore, as CIP2A is one of the most potent PP2A inhibitor with clinical relevance, inhibition of CIP2A protein expression or activity could constitute an effective cancer therapy strategy. However, before drug development can begin, CIP2A mRNA and/or protein variants should first be identified. At the moment virtually nothing is known about variant forms of CIP2A. In addition, a previous study had suggested a link between CIP2A and a stem cell marker OCT4, but the functional and/or regulatory relationship between these proteins was not clear.

The specific aims of this thesis were:

1. To study the nature of the regulatory relationship between OCT4 and CIP2A
2. To do a systematic analysis of PP2A inhibitor protein mRNA expression in AML and to determine the clinical relevance of these findings
3. To identify novel CIP2A variants and to determine their clinical relevance in cancer

4 Materials and Methods

4.1 Cell culture and transfections

4.1.1 Cell lines and cell culture (I–III)

All cell lines were cultured at 37ºC in a 5% CO_2 humidified atmosphere and were routinely tested for mycoplasma contamination. All cell culture media were supplemented with heat inactivated fetal bovine serum (FBS, Gibco, Thermo Fisher Scientific, Waltham, MA, USA), 2 mM L-glutamine (Sigma-Aldrich), 50 units/ml penicillin (Sigma-Aldrich) and 50 mg/ml streptomycin (Sigma-Aldrich) if not otherwise indicated.

Acute myeloid leukemia (AML) cell lines KG-1 (ACC-14), HL-60 (ACC-3), MOLM-14 (ACC-777) and KASUMI-1 (ACC-220) were obtained from Leibniz-Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). KG-1, HL-60, MOLM-14 and KASUMI-1 cells were maintained in RPMI-1640 medium (R5886, Sigma-Aldrich, Saint Louis, MO, USA) supplemented with 10% (KG-1, HL-60) or 20% (MOLM-14, KASUMI-1) FBS. AML cell lines F36P, HEL, TF-1, EOL-1 and MOLM-13, and chronic myeloid leukemia (CML) cell lines K562, KU812 and MEG-01 were obtained from Maria Odero's laboratory (University of Navarra, Pamplona, Spain). HEL, EOL-1, K562 and MEG-01 cells were maintained in RPMI-1640 supplemented with 10% FBS. MOLM-13, KU812, F36P and TF-1 were maintained in RPMI-1640 with 20% FBS, and F36P and TF-1 media additionally included 10 ng/mL GM-CSF. Breast cancer MCF7 and MDA-MB-231, and cervical cancer HeLa, cell lines were purchased from ATCC (Manassas, VA, USA). MCF7 cells were cultured in RPMI supplemented with 10% FBS, 10µg/ml insulin and 100µM NEAA. MDA-MB-231 and HeLa cells were maintained in DMEM (Dulbecco's modified Eagle's medium, Sigma-Aldrich) supplemented with 10% FBS. Glioblastoma cell lines BT5, BT3 and BT3CD133 were acquired from Pirjo Laakkonen's laboratory (University of Helsinki, Helsinki, Finland). BT5, BT3 and BT3CD133 cells were cultivated as adherent cultures in serum-free neural stem cell (NSC) media containing an equal mix of DMEM/F12 (1:1) with glutamax (31331-093, Fisher Scientific) and neurobasal (21103-049, Fisher Scientific) media with 1X B-27 (12587001, Fisher Scientific) 1× N-2

(17502001, Fisher Scientific) supplements, 1% penicillin-streptomycin (P4333, Sigma-Aldrich), bFGF (10 ng/ml, 100-18B, Peprotech) and EGF (10 ng/ml, AF-100-15, Peprotech) (NSC+/+ medium). T98G cells (VTT Technical Research Centre, Turku, Finland in 2010) were cultured in Eagle MEM (Sigma-Aldrich) supplemented with 10% FBS. Human head and neck squamous cell carcinoma (HNSCC) cell lines UT-SCC7, UT-SCC2, UU-SCC9, UT-SCC105, UT-SCC111, UT-SCC115, UT-SCC118, UT-SCC12A, UT-SCC59A and UT-SCC91 were established at the time of operation from HNSCCs (Lansford et al. 1999) and when in culture, cells were maintained in DMEM medium supplemented with 10% FBS. Testicular cancer cell lines Tcam2 and Tera1 were maintained in RPMI-1640 supplemented with 10% FBS.

4.1.2 Transient transfections with siRNA (I, II)

Cells were silenced with small interfering RNA (siRNA) when the confluency was approximately 30 %. siRNA silencing was performed using 20-250 nM siRNA and Lipofectamine 2000 reagent (Thermo Fisher Scientific) diluted to Opti-MEM (Gibco) according to manufacturer's instructions for adherent cell lines. For suspension cell lines, optimized transfection programs for each cell line were performed with Nucleofector II Device (Lonza, Basel, Switzerland). 24h post-silencing, transfection solution was changed to full cell medium. Cells were used for experiments 1–5 days after silencing.

Two stranded siRNA oligonucleotides used in the original publications I, II: siGENOME ARPP19 siRNA (D-015338-03, Dharmacon, Lafayette, CO, USA), CIP2A-1 siRNA (5'-CUGUGGUUGUGUUUGCACU(dTdT)-3'), CIP2A-2 siRNA (5'-ACCAUUGAUAUCCUUAGAA(dTdT)-3'), OCT4-1 siRNA (5'-GGAGAAGCUGGAGCAAAAC(dTdT)-3'), OCT4-2 siRNA (5'-GCCGGGCUGGGUUGAUCCU(dTdT)-3'), non-silencing control SCR-1 siRNA (5'-CGUACGCGGAAUACUUCGA-(dTdT)-3') and non-silencing control SCR-2 (5'-CCUACAUCCCGAUCGAUGAUG(dTdT)-3').

4.1.3 Generation of stable cell lines with retroviral infections and cell viability assay (II)

Small hairpin RNA (shRNA) constructs were ordered as lentiviral particles from the Biomedicum Functional Genomics Unit ((FuGU), University of Helsinki, Finland) TRC1 library. ARPP19 shRNAs were TRCN0000158847 and TRCN0000160408. Control shSCR was SHC002 (Sigma). To establish the stable cell line, the ARPP19-RNAi lentivirus was transfected into HL-60 and KG1 cells with several different amounts of infectious virus. Twenty-four hours after transduction, spinoculation was

performed and selection was done with puromycin at the 72 h time point. ARPP19 expression was determined through Western blot analysis and RQ-PCR.

Differences in cell viability of shARPP19 transduced cell lines compared to control shRNA cell lines were measured with CellTiter-Glo® luminescent assay (Promega, Fitchburg, WI, USA) at 24 h, 48 h, 72 h, 96 h or 120 h after plating the cells. Results were derived from the average of at least three independent experiments.

4.1.4 Promoter assay with plasmid transfections (I)

Tcam2-cells were double transfected using Surefect transfection reagent (Nunclon Surface, Thermo Fisher Scientific) according to manufacturer's instructions. Cells were transfected simultaneously with CIP2A-promoter construct (-1802 bp upstream (Khanna et al. 2011)), renilla plasmid and siRNA (either non silencing control SCR-1 or siOCT4-1). 200 ng of promoter construct, 10 ng of renilla plasmid and 2 pmol of siRNA were transfected per 96-well plate well. Transfections with -1802 bp, -865 bp and -1802ΔCIP2ALuc CIP2A promoter constructs were done as described above only without siRNAs. -1802ΔCIP2ALuc construct was produced by GenScript mutagenesis service from -1802CIP2ALuc promoter construct, and the resulting promoter sequence was validated by DNA sequencing. 72h post-transfection, CIP2A promoter activity was measured using Dual-Glo luciferase Assay system (E2940, Promega) according to manufacturer's instructions. Luminescence was measured with Victor-multilabel counter 1420 (PerkinElmer Inc, Waltham, MA, USA). Results were derived from the average of at least three independent experiments and two technical replicates.

4.1.5 Mouse cell lines and *in vitro* studies (I)

4.1.5.1 Derivation of mouse embryonic stem cells from blastocysts (I)

Mouse embryonic stem cells (mESCs) were isolated from CIP2A homozygote and wild type blastocysts as described by Bryja and coworkers (Bryja, Bonilla, and Arenas 2006). Time-mated females were killed at E3.5, and the blastocysts were flushed out of the uterine horn. Blastocysts were plated to dishes containing mitotically inactivated feeder cells (mouse embryonic fibroblasts, MEFs). Blastocysts were allowed to attach to MEFs and grow in ES medium containing knockout serum replacement (SR-ES medium). SR-ES medium contained: Knockout DMEM supplemented with 20% Knockout SR (Gibco), penicillin (100 U/ml)/streptomycin (100 g/ml) (Gibco), 2 mM L-glutamine (Gibco), 1 X minimal essential medium nonessential amino acids (Gibco), 100 μM β-mercaptoethanol and

recombinant mouse leukemia inhibitory factor (1,000 U/ml of ESGRO, Chemicon International, Temecula, CA). The blastocysts and mESCs derived from the inner cell mass of blastocysts were allowed to grow alternately in SR-ES and FCS-ES medium. In FCS-ES medium SR was replaced by 20% fetal calf serum (FCS). The cells were grown in FCS-ES for one day after each trypsinization to allow greater trophic support, whereas SR-ES medium was used to support the selective propagation of ESCs between the trypsinizations.

4.1.5.2 ZHBTc4-mESC *in vitro* studies (I)

ZHBTc4 undifferentiated mouse embryonic stem cells (Niwa, Miyazaki, and Smith 2000)) were kindly provided by Dr. Hitoshi Niwa (Center For Developmental Biology, Laboratory for Pluripotent Cell Studies, Kobe, Japan). ZHBTc4-mES cells lack functional endogenous *Oct4* alleles and harbor a regulatable *Oct4* transgene that can be conditionally downregulated by doxycycline treatment. ZHBTc4-mES cells were kept in undifferentiated state by culturing them on a feeder layer of mitomycin C-inactivated mouse embryonic fibroblasts with basic ES cell medium. The cells were passaged every two-three days and ES cell medium was exchanged daily. To study the effect of OCT4-mediated differentiation of mES cells, the cells were plated on to 0.1% gelatin-coated culture dishes and cultured with or without 1 μg/ml doxycycline. Cells were harvested 6, 12, 24, 48 and 72 hours by scraping off the cells, pelleting them by centrifugation and snap-freezing them in liquid nitrogen. Results were derived from the average of three independent experiments.

4.2 Gene and protein expression measurements

4.2.1 RNA isolation and cDNA synthesis (I–III)

Total RNA was isolated from cell lines, extracted from bone marrow (BM; AML) or peripheral blood (CML) mononuclear cell pellets or seminiferous tubules of mouse testis. Total RNA from cell lines was extracted with NucleoSpin RNA II kit (MACHEREY-NAGEL), from AML patient sample mononuclear cells with E.Z.N.A.® Total RNA Kit I (Omega Bio-Tek Inc, Norcross, GA, USA), from CML patient sample mononuclear cells using RNeasy mini kit (Qiagen, Hilden, Germany) and from seminiferous tubules of mouse testis using Trisure reagent (Bioline, London, UK) according to the manufacturer's instructions. If not included in the RNA extraction procedure by the manufacturer, additional DNase I treatment was conducted to remove traces of genomic DNA from the samples. After isolation, RNA concentration and RNA purity were measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

In the original publication I, DyNAmo SYBR Green 2-step qRT-PCR Kit (Finnzymes, Espoo, Finland) was used for cDNA synthesis of the mouse samples. In the original publications II and III, cDNA of cell line samples was synthesized (with 1 µg of total RNA as a starting material) using M-MLV Reverse Transcriptase (M3682, Promega) RNase (H-) (M3682, Promega), random primers (C1181, Promega), RNasin Ribonuclease Inhibitor (N2511, Promega) and dNTP mix (#R0192, Thermo Fisher Scientific). cDNA of patient samples was synthesized using SuperScript III Reverse Transcriptase (18080093, Invitrogen), random primers (C1181, Promega), RiboLock(tm) Ribonuclease Inhibitor (#EO0381, Thermo Fisher scientific) and dNTP-mix (BIO-39028, Bioline). RT-reactions were performed according to each enzyme's manufacturer's instructions.

4.2.2 Quantitative real-time polymerase chain reaction (I–III)

Primers for each gene specific assays, when possible, were designed to be located to different exonic sequences to avoid amplification of genomic DNA. The primer concentration in each reaction was 300 nM and probe concentration 200 nM. The specificity of quantitative real-time PCR (RQ-PCR) reactions was verified by agarose gel electrophoresis and melting curve analysis (DyNAmo Flash SYBR Green qPCR Kit, Thermo Fisher Scientific); a single band of the expected size and a single peak, respectively, were required. The amplification efficiency for each target was also assessed. Results were derived from the average of at least two independent experiments and two technical replicates. The primer and probe sequences used in this study for RQ-PCR analysis of human and mouse genes are listed in Table 3 and Table 4.

Amplification of target cDNAs in human samples were performed using KAPA PROBE FAST RQ-PCR Kit (Kapa Biosystems, Wilmington, MA, USA) and 7900 HT Fast Real-Time PCR System (Thermo Fisher Scientific) according to the manufacturers' instructions. RQ-PCR was executed under the following conditions: 95°C for 10 min followed by 45 cycles of 95°C for 15 s and 60°C for 1min. Relative gene expression data were normalized to the expression level of endogenous house-keeping genes glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and beta-actin using the 2^-ΔΔC(t) method with SDS software (version 2.4.1, Applied Biosystems, Foster City, CA, USA) or with Thermo Fisher Cloud Real-time qPCR Relative Quantification application.

For mouse samples in the original publication I, amplification of target cDNAs were performed using CFX96 real-time PCR detection system device (Bio-Rad Laboratories Inc., Hercules, CA, USA) and the DyNAmo Flash SYBR green qPCR kit (Thermo Fisher Scientific) according to the manufacturers' instructions. Quantitative real-time PCR was executed under the following conditions: 95°C for

7 min followed by 40 cycles of 94°C for 1 s and 55–64°C (depending on the primer pair; see table XX) for 15s. Relative gene expression data were normalized to expression levels of endogenous house-keeping genes cyclophilin A (*Ppia*) and ribosomal protein L19 (*Rpl19*) using the 2^-ΔΔC(t) method.

To estimate the degree of overexpression in the AML and CML patient cohorts, the expression of each gene was normalized to the expression level in a commercial normal BM control sample (pooled from 56 males and females, 636591, lot 1002008, Clontech Laboratories, Fremont, CA, USA). In CML cohort2 a pool of cDNA from 4 normal individuals was used as calibrator and all the samples were normalized to GAPDH as an endogenous control.

Table 3. Human primer and probe sequences used in this study for RQ-PCR analysis.

Target	Forward primer Reverse primer	Probe	Used in
ARPP19	cagagggagcactatgtctgc gcttttaattttgcttcttctgct	aggagcag*	II, III
β-ACTIN	tcacccacacrgtgcccatctacgc cagcggaaccgctcattgccaatgg	atgccctcccccatgccatcctgcgt	I, II, III
CIP2A (e13)	cagtctggactgagaatattattgga ggcattgtttgctgctatacttt	tccactgc*	III
CIP2A (e20)	gaacagataagaaaagagttgagcatt cgaccttctaattgtgcctttt	cttcctcc*	I, II, III
EVI1 (MECOM)	agtgccctggagatgagttg tttgaggctatctgtgaagtgc	ccccagtgaggtataaagagga	II, III
GAPDH	acccactcctccaccttga ttgctgtagccaaattcgttgt	acgaccactttgtcaagctcatttcctggt	I, II, III
*GAPDH***	acccactcctccaccttga ttgctgtagccaaattcgttgt	acgaccactttgtcaagctcatttcctggt**	III
*NANOG****	gatttgtgggcctgaagaaa aagtgggttgtttgcctttg	Annealing temperature 57.5ºC	I
NOCIVA#1	cagcctcactgaacatggaa cagtcaaaatggtgggaagg	P34*	III
NOCIVA#2	aaaagtggtctgtaaagctgctc ttcatgtcagattaccagtttatgc	P49*	III
*NOCIVA#3***	atgccaagacacagtcaaaatg cctgcttgcataaactggtaatc	cagaggcagaggataa**	III
OCT4	agcaaaacccggaggagt ccacatcggcctgtgtatatc	P35*	I
PPME1	acaggtttgcagaacccatc ggacagcaggtcactaacagc	tccagtgt*	II, III
TIPRL	catgatgatccacggcttc tcagggagagatggcatatgta	ggccctgg*	II, III
WT1	gggcgtgtgaccgtagct cgctattcgcaatcagggtta	agcacggtcaccttcgacggg	II, III

* Roche Universal ProbeLibrary (UPL) probe; ** MGB Probe 6-FAM- seq -MGB-Eclipse®3`; *** SYBR Green assay.

Table 4. Mouse primer and probe sequences used in this study for RQ-PCR analysis. All assays SYBR Green assays.

Target	Forward primer Reverse primer	Annealing temperature	Used in
Cd9	tgcagtgcttgctattggac ggcgaatatcaccaagagga	56ºC	I
Cip2a	gcgccatgtactcagtcaga aggaagcagaagggtcacaa	57ºC	I
c-Kit	atcccgactttgtcagatgg aaggccaaccaggaaaagtt	56ºC	I
Oct4	cacgagtggaaagcaactca agatggtggtctggctgaac	64ºC	I
Plzf	aacggttcctggacagttg cccacacagcagacagaaga	59ºC	I
Ppia	catcctaaagcatacaggtcctg tccatggcttccacaatgtt	63ºC	I
Rpl19	ggacagagtcttgatgatctc ctgaaggtcaaagggaatgtg	55ºC	I
Stra8	atgcaatgttgctgaagtgc ggaagcagcctttctcaatg	57ºC	I

4.2.3 Western blot assay (I–III)

Cells were first resuspended in Triton-X-100 lysis buffer (TXLB; 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% Triton-X-100, 5% glycerol, 1% SDS) or SDS buffer (150 mM Tris-HCl (pH 6.8), 15% β-mercaptoethanol, 30% Glycerol, 1.2% SDS) supplemented with protease and phosphatase inhibitors (Thermo Fisher Scientific). Cells were lysed on ice for 5 minutes (0.5min ON/0.5min OFF with full power) using a Bioruptor® sonicator (Diagenode SA, Seraing, Belgium). Protein extracts were centrifuged at high speed (10,000 g for 5 min) and the supernatant was transferred to a new tube to remove cell debris. Protein concentration was estimated using a DC Protein assay (Bio-Rad Laboratories, Hercules, CA, USA). Sample buffer was added and samples were boiled for 5 min at 95°C. Equal amounts of protein were loaded onto acrylamide gels (4-20% Mini-PROTEAN®TGX Gels, Bio-Rad Laboratories). Protein extracts were separated on SDS-PAGE under denaturing conditions and transferred to PVDF membrane (Bio-Rad Laboratories) by semi-dry turbo blot (Bio-Rad Laboratories). Membranes were blocked with 5%-milk-TBST (Tris-buffered saline and 0.1% Tween 20) for 1 hour at RT. Primary antibodies were diluted in 5%-milk-TBST and incubated with membranes overnight at 4°C. After primary antibody incubation, membranes were washed three times with TBST for 5 min at RT. Amersham ECL HRP-linked secondary antibodies (GE Healthcare) were diluted 1:5000 in 5%-milk-TBST and incubated with membranes at RT for 1 h. Amersham ECL Plus™ Western blotting reagent (GE Healthcare) was added to the membrane and signals were detected using enchanced chemiluminescence on developed film.

Band intensity was determined using ImageJ software (National Institutes of Health (NIH), Bethesda, MD, USA). Results were derived from the average of at least three independent experiments. Primary and secondary antibodies used in the study are presented in Table 5 and Table 6.

Table 5. Primary antibodies used in the study. Application abbreviations: IF, immunofluorescence; BA, binding assay; WB, Western blotting. Manufacturer abbreviations: SCBT, Santa Cruz Biotechnology; TFS, Thermo Fisher Scientific.

Primary antibody	Type	Manufacturer	Catalog no.	Application	Used in
Alexa Fluor® 647 CD24	Mouse monoclonal	SCBT	561644	Cell sorting	I
ARPP19	Rabbit polyclonal	Proteintech Group	11678-1-AP	WB	II
PP2A-B56α	Mouse monoclonal	SCBT	sc-136045	WB, BA	III
CD44	Rat monoclonal	in house	clone 9B5	Cell sorting	I
CDK1 (Cdc2 p34)	Mouse monoclonal	SCBT	sc-51578	WB	II
CIP2A	Mouse monoclonal	SCBT	sc-80659	WB, IF	I, II, III
CIP2A	Rabbit polyclonal	Soo Hoo et al. 2002	-	IHC	I
CYCLIN D1	Mouse monoclonal	SCBT	sc-450	WB	II
C23 (nucleolin)	Mouse monoclonal	SCBT	sc-8031	IF	III
GAPDH	Mouse monoclonal	Hytest	5G4-6C5	WB	I, II, III
GST	Rabbit polyclonal	TFS	CAB4169	WB, BA	III
KI-67	Mouse monoclonal	DAKO	M7240	IHC	I
MYC	Mouse monoclonal	SCBT	sc-40	WB, IF	I, II
MYC	Mouse monoclonal	Nordic Biosite	LS-C821368	IHC	I
MYC phospho S62	Mouse monoclonal	Abcam	ab78318	WB	I
NOCIVA #1	Rabbit monoclonal	BioGenes GmbH	-	WB, IF	III
NOCIVA #2	Rabbit monoclonal	BioGenes GmbH	-	WB, IF	III
OCT4	Mouse monoclonal	SCBT	sc-5279	WB, IHC	I
OCT4	Rabbit polyclonal	SCBT	sc-9081	WB	I
V5	Mouse monoclonal	TFS	E10/V4RR	WB, BA	III
β-ACTIN	Mouse monoclonal	Sigma-Aldrich	AC-74	WB	I, II, III

Table 6. Secondary antibodies used in the study. Application abbreviations: IF, immunofluorescence; WB, Western blotting. Manufacturer abbreviations: TFS, Thermo Fisher Scientific.

Secondary antibody	Type	Manu-facturer	Catalog no.	Application	Used in
Alexa Fluor® 488 anti-mouse IgG	Goat polyclonal	TFS	A32723	IF	III
Alexa Fluor® 488 anti-rabbit IgG	Goat polyclonal	TFS	A32731	IF	III
Alexa Fluor® 488 anti-rat IgG	Goat polyclonal	TFS	A-11006	Cell sorting	I
Alexa Fluor® 568 anti-mouse IgG	Goat polyclonal	TFS	A-11031	IF	III
Alexa Fluor® 568 anti-rabbit IgG	Goat polyclonal	TFS	A-11036	IF	III
anti-mouse-HRP	Goat polyclonal	DAKO	P0447	WB	I, II, III
anti-rabbit-HRP	Goat polyclonal	DAKO	P0399	WB	I, II, III
DAKO EnVision™ anti-mouse-HRP	Goat polyclonal	DAKO	K4000	IHC	I
DAKO EnVision™ anti-rabbit-HRP	Goat polyclonal	DAKO	K4002	IHC	I

4.3 Flow cytometry-based assays

4.3.1 Cell sorting (I)

UT-SCC2, -7 and -9 cells were harvested with 0.01% Trypsin-EDTA and washed twice with cold buffer (D-PBS, 2% FCS, 0.01% sodium azide). Primary antibodies (anti-human CD44 (clone 9B5) rat monoclonal antibody was a kind gift from Professor Marko Salmi (Turku, Finland), anti-human CD24 (clone ML5) Alexa Fluor® 647 mouse monoclonal antibody (BD Biosciences)) were added at dilution of 1:100 and incubated for 1 hour at +4°C, after which cells were washed. Secondary antibody (Alexa Fluor® 488 Goat Anti-Rat IgG (Life Technologies)) was added at dilution of 1:400 and incubated for 1 hour at +4°C. The cells were washed and sorted with BD FACSAria™ III cell sorter (BD Biosciences, Franklin Lakes, NJ, USA). After sorting the cells were lysed with TXLB buffer. Three independent sorting experiments were conducted for each cell line.

4.3.2 Cell cycle analysis (II)

Cells were incubated 24-48 h to reach 50-60 % confluence, collected by trypzination, washed twice with cold buffer (D-PBS, 2% FCS, 0.01% sodium azide) and fixed with 70% ethanol at -20°C for 24 h. The fixing solution was removed by centrifugation and samples were washed with PBS. Ribonuclease (50 µl of a 100 µg/ml stock of RNase) treatment was performed to remove RNA from the samples.

Staining was done at RT with propidium iodide (10 µg/ml) in PBS with 0,1% Tween. Cell cycle analysis was performed using flow cytometer BD FACSCalibur (BD Biosciences) and analyzed using Flowing Software 2.5. The median fluorescence intensity was measured from 5000–20000 events. Viable single cells were gated by forward scatter (FSC) and side scatter (SSC) dot plot. Results were derived from the average of three independent experiments.

4.4 NOCIVA discovery PCR (III)

4.4.1 Rapid amplification of cDNA ends (III)

Extracted total RNA with Macherey-Nagel NucleoSpin RNA II -kit was used as a template for rapid amplification of cDNA ends (RACE) experiments. For both 3' and 5' end cDNA amplification, Invitrogen's (Carlsbad, CA, USA) 3'RACE (catalog no. 183743-019) and 5'RACE (catalog no. 18374-058) kits were used according to manufactures protocols. Multiple gene specific primers (GSPs) were designed and used for proper CIP2A related amplification of cDNA ends. Amplified sequences were run on agarose gels (percentage linked to the predicted fragment sizes), cut, DNA extracted (NucleoSpin® Gel and PCR Clean-up, MACHEREY-NAGEL, Düren, Germany) and DNA sequenced for analysis. Sequencing primers were designed to be downstream or upstream of the GSPs. Main GSP and sequencing primer sequences are listed in Table 7.

4.4.2 Validation polymerase chain reaction (III)

PCR for NOCIVA validation analysis was performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) and PCR-reactions, including temperatures and volumes, were performed according to enzyme's manufacturer's instructions. Main primer sequences used for PCR analysis are listed in supplemental Table 7.

Table 7. Main primers used in the study. Application abbreviations: PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends.

Target	Primer sequence	Application	Used in
CIP2A exon13 (GSP1)	tacttcaggacccacgtttgattact	3'RACE	III
CIP2A exon12 (GSP2)	cattggttcctggtatggaagtaagc	3'RACE	III
CIP2A exon6 (GSP3)	cgataaaaagatttcaca	5'RACE	III
CIP2A exon7-8 (GSP4)	gctcatatctggtg	5'RACE	III
CIP2A exon13 F2	ctgctccactgccagattt	PCR	III
CIP2A exon13 F5	tcaggacccacgtttgattac	PCR	III
R1_NOCIVA	gcagaggataagacttccatgtt	PCR/DNAseq	III
R2_NOCIVA	atagagcagctttacagaccac	PCR/DNAseq	III
R3_NOCIVA	tgtaacttgaaaagcttcctgtgta	PCR/DNAseq	III
F1_CIP2A exon1	cctgaattcccatggactccactgcct	PCR/DNAseq	III
F2_CIP2A exon1	atggactccactgcct	PCR/DNAseq	III

4.5 Protein measurements

4.5.1 Protein expression and purification (III)

The truncated domains of human CIP2A (1-560, 1-330, 561-905) and full length NOCIVA were cloned in pGEX vector (GE Healthcare, Chicago, IL, USA), which produces proteins as thrombin-cleavable amino-terminal glutathione S-transferase (GST)-fusion proteins, for expression in Escherichia coli, and was verified by sequencing. BL21 (DE3) cells (Merck, Darmstadt, Germany) cells were used for overexpression. The overnight bacterial culture was inoculated in LB media and incubated at 37°C until OD600 reached 0.6-0.9. Expression was induced with 0.2 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for about 4 hours at 23°C. The bacterial pellets were collected by centrifugation at 6,000 g at 4°C and stored at -20°C until purification. Cells were lysed by sonication on ice, in a buffer consisting of 200 mM Tris pH8, 500 mM NaCl, 2 mM dithiothreitol (DTT, Sigma-Aldrich, Saint Louis, MO, USA), 0.5% Tx-100, lysozyme (20 mg/ 150 mL, Calbiochem 4403-1GM), and 1 x Pierce Protease Inhibitor Mini Tablets, EDTA-Free (Thermo Scientific, A32955). The cleared lysate was incubated with 750 µl glutathione agarose slurry (1:1 diluted with lysis buffer, Glutathione Sepharose 4B, 17-0756-01, GE Healthcare) for about 3 hours at 4°C, with gentle rotation. Pelleted beads were washed extensively with washing buffer (same composition as lysis buffer, but without lysozyme), and then eluted using elution buffer: 100 mM Tris pH8, 200 mM

NaCl, 5 mM DTT, 0.1% Tx-100 and 20 mM Glutathione (L-Glutathione Reduced; Sigma-Aldrich, G4251-5G). The sample was dialysed using Snakeskin MWCO 10k tubing (Thermo Scientific, 88243) into a buffer containing 20 mM Tris pH8, 150 mM NaCl, 2 mM DTT, 0.05% Tx-100 and 10% glycerol. Next, the pulled fractions were concentrated using MWCO tubing (Merck Millipore), and concentration was determined by Coomassie staining (PageBlue Protein Staining Solution, Thermo Scientific, 24630), using GST alone as internal standard.

4.5.2 Binding assay (III)

In binding assays, all the purified recombinant proteins were used at 10 pmol. Protein samples were diluted in reaction buffer (50 mM Tris HCl pH 7.5, 150 mM NaCl, 2 mM DTT, 0.2% Igepal, 10% glycerol) and incubated for 1 hour at 37°. The reaction volume was 150 µl. Next, 5 µl input sample was withdrawn prior to adding 5 µl glutathione agarose (Glutathione Sepharose 4B,17-0756-01, GE Healthcare) (diluted 4 x) to each sample, and precipitated complexes formed for 1 hour at RT, by incubating samples with moderate rotation. The beads were washed by adding 250 µl of reaction buffer for a total of four buffer exchanges and for 1 hour at 4°C, with moderate rotation. The bound complexes were eluted off the beads by adding 30 µl 2 x SDS-PAGE sample buffer, incubating for 10 min at 95°C, then recovering the eluted proteins by centrifuging at 3,000g for 1 min. The eluted materials were resolved on 4-20% SDS-PAGE (Mini-Protean TGX Gels, Bio-Rad), transferred on PVDF membrane and blotted as indicated for B56α (1:5000, sc136045, Santa Cruz Biotechnology), V5 (1:5000, E10/V4RR, Thermo Fisher Scientific) and GST (1:10000, CAB4169, Thermo Fisher Scientific). Results were derived from the average of three independent experiments.

4.5.3 Immunocytochemistry and imaging (III)

For immunocytochemistry, cells were grown on coverslips, fixed with 4% PFA in PBS for 10 min at RT and permeabilizied with 0.1% TritonX-100 in PBS for 10 min at RT. Coverslips were incubated with primary antibodies (1:100, CIP2A, NOCIVA, nucleolin) in PBS for 1 hour at RT, and secondary antibodies (1:300, 488 and 555 Alexa Fluor-conjugated, Thermo Fisher Scientific) in PBS for 1 hour at RT. Cell nuclei were stained with 4'6-diamidino-2-phenylindole (DAPI, 1:10000, Life Technologies). Coverslips were mounted with Mowiol 4-88 (Sigma-Aldrich) mounting medium and imaging was performed using LSM780 (Carl Zeiss) confocal microscope with C-Apochromat 40x/1.20 W Korr M27 objective at the Cell Imaging Core at Turku Bioscience.

4.6 *In vivo* methods

All mice used for experiments were housed in plastic cages (Tecniplast, Buguggiate, Italy) in a climate-controlled room at the Animal Centre of Turku University. Aspen chips (Tapvei Co., Kaavi, Finland) were used as bedding material. Animals were maintained on a 12 h light/12 h dark cycle and they had free access to tap water and standard laboratory animal feed (Commercial RM3 (E) SQC, Special Diet Service, Witham, UK). The Regional State Administrative Agency for Southern Finland had ethically assessed and authorized all animal work protocols (ESLH-2007-08517). All animal experiments were conducted in accordance with the guidelines of the Provincial Government of Southern Finland and handled in accordance with the institutional animal care policies of the University of Turku.

4.6.1 *In vivo* X-irradiation (I)

Eight weeks old male C57BL/6 mice from The Jackson Laboratory were anaesthetized with intraperitoneal injection of 2.5% Avertin (Aldrich Chemical Co., Milwaukee, WI, USA), and the testes were locally irradiated (with a water-equivalent build-up layer, focus-target distance 100 cm, field size 4 x 10 cm, dose rate 3 Gy/min) with 3-4 Gy using 6 MV X-rays produced by a Clinac 600C linear accelerator (Varian, Palo Alto, CA, USA). Radiation dose in testis was determined mathematically using a computer tomography-based Eclipse planning system (Varian). To avoid systemic side-effects, mouse testes were X-irradiated with 4 Gy under CT-scan guidance (I, Fig. 1B). The mice were sacrificed by neck dislocation under CO_2 anesthesia 6, 17, 24, 48, 72, 96 or 144 hours after X-irradiation and their testes were dissected and decapsulated. Seminiferous tubules were snap-frozen in liquid nitrogen and used in RNA analyses. Control mice were subjected to the same treatment omitting the X-irradiation. Genes that did not show changes in expression were considered to be expressed in radioresistant spermatogonial stem cells.

4.6.2 Subcutaneous tumor xenografts in nude mice (I)

Previously established HNSCC cell lines (Pekkola-Heino et al. 1991), UT-SCC-14 (originating from a persistent T3N1M0 Gr 2 cancer of the mobile tongue) and UT-SCC-50 (established from a recurrent T2N0M0 Gr 3 glottic laryngeal tumor), were selected for *in vivo* tumor formation experiment. Two million cells were injected subcutaneously in 100 µl (50% PBS, 50% Matrigel) in the flank of 6-8 weeks old immunocompromised Hsd:athymic nude mice. Altogether six mice were injected (3+3) and the size of the palpable tumors was measured with a caliper every third day for five weeks. After this, all mice were sacrificed and the diameter of the final tumors were measured.

4.6.3 Immunohistochemistry and tissue samples (I)

Formalin-fixed and paraffin-embedded sections of mouse and human tissue samples were cut to 6 μm thin sections, deparaffinized and rehydrated. The endogenous peroxidase activity in tissues was blocked using 3% hydrogen peroxide. For immunohistochemistry, heat-mediated antigen retrieval was carried out in Tris-EDTA-buffer (pH 9.0, 10 mM Tris, 1 mM EDTA) by heating sections in microwave oven (4 min, 850 W followed by 15 min, 150 W). Sections were then blocked with 3% BSA PBS for 10 min at RT and the slides were rinsed with Tris-HCl (pH 7.4). Next, slides were incubated overnight at 4°C with primary antibodies against CIP2A (1:10000, (Soo Hoo, Zhang, and Chan 2002), OCT4 (1:200, sc-5279, Santa Cruz Biotechnology), KI-67 (1:5000, M7240, DAKO), or MYC (1:200, LS-C821368, Nordic Biosite). Control slides were incubated with normal nonimmunized appropriate animal serum. After washes, the samples were incubated for 30 min at RT with appropriate secondary antibody (DAKO EnVision anti-rabbit or anti-mouse) and washed again. Then, DAB+ solution (DAKO, K3468) was added for 10 min at RT followed by washing. After counterstain with Mayer's HTX, slides were dehydrated, cleared in xylene and mounted in Pertex. Stained samples were imaged with Pannoramic P1000 slide scanner (3DHISTECH Ltd) and analyzed visually. The usage of human tissue samples was approved by the Finnish national authority for medicolegal affairs (Dnro 889/04/047/08) and regional ethics committee of University of Turku (Dnro 146/2007).

4.7 Patient samples and end point definitions

4.7.1 Acute myeloid leukemia patient cohorts (II, III)

4.7.1.1 AML study cohort1 (II, III)

Consecutive bone marrow samples were collected between January 2000 and July 2010, a total of 80 patients aged 18–65 diagnosed with *de novo* or secondary AML at Turku University Hospital (TYKS). Patients with acute promyelocytic leukemia (t(15;17)(q22;q12)) were excluded from this cohort. Patient characteristics are presented in the original publication II, Table S1. Their median age was 50 years (Q_1 = 38.8, Q_3 = 58.0), median overall survival was 5.4 years (95% CI, 2.8 to 7.9) and median follow-up time was 5.4 years (range 6 days–16 years). The ELN risk classification, based on cytogenetic and molecular findings, was used as risk stratification (II, Table S2). Most patients (76) were enrolled in the Finnish Leukemia Study Group prospective protocols (II, Table S3). In total, 32 patients were treated according to AML92 and 44 according to AML2003 protocol.

Treatment of four patients was significantly modified due to patient-related reasons. Although patients were treated with different schedules, all received regimens based on anthracycline and high-dose cytarabine as induction therapy. High-dose cytarabine and allogenic stem cell transplantation when possible, were used as consolidation therapy. No significant differences were found between the relapse or overall survival rates of patients on the AML92 and the AML2003 treatment. Informed consent was obtained from all patients and the local Ethical Review Board of TYKS approved the study protocol. No missing data imputation was performed.

4.7.1.2 AML study cohort2 (II)

Bone marrow samples from 48 AML patients, including nine AML patients with supplementary follow-up samples at first remission and/or at relapse, were analyzed from the Finnish Hematology Registry and Clinical Biobank (FHRB) collection. Patient characteristics for the nine patients are presented in the original publication II, Table S4. All 48 patients had received intensive chemotherapy as an induction therapy and achieved CR. Additional follow-up samples at remission were available from four patients and at relapse from eight patients. Samples were collected from Finnish university hospitals and other hematological units between December 2011 and January 2017. Median age for the nine patients was 59.8 years (Q_1 = 50.7, Q_3 = 68.8), median overall survival was 1.7 years (95% CI, 1.3 to 3.9) and median follow-up time was 1.7 years (range 1–4.5 years). FHRB is authorized by the Finnish National Supervisory Authority for Welfare and Health (Valvira) and has been approved by the Finnish National Medical Ethics Committee. All patients signed an informed consent prior to biobanking.

4.7.1.3 TCGA LAML dataset (II)

RNA sequencing dataset available from The Cancer Genome Atlas for AML patient cohort (TCGA LAML (Ley et al. 2013), survival data available for n = 160, exon expression IlluminaHiSeq) was analyzed for the correlation between OS and ARPP19 gene expression using UCSC Xena Browser (Goldman et al. 2018).

4.7.2 Chronic myeloid leukemia patient cohorts (III)

4.7.2.1 CML study cohort1 (III)

CML patient cohort1 comprised of 35 newly diagnosed CP CML patients from the University of Liverpool CML biobank. One patient was excluded from survival analysis as there were no follow up data. Twenty patients received imatinib and 14

received a second generation TKI, either dasatinib or nilotinib, as a first-line therapy. Their median age was 53.5 years (Q_1 = 42.3, Q_3 = 62.0), the median follow-up time was 32.5 months (range 9-75 months) and median event free survival was 30.9 months (95% CI, 24.1 to 39.4). All patients had signed an informed consent prior to biobanking.

4.7.2.2 CML study cohort2 (III)

CML study cohort2 included 159 newly diagnosed CP CML patients from the UK-wide SPIRIT2 clinical trial. In the SPIRIT2 trial, 814 newly diagnosed chronic phase patients were randomly allocated 1:1 to either imatinib 400mg or dasatinib 100mg each once daily between 2008 and 2013. Follow-up was monthly for 3 months, 3-monthly until 12 months, then 6-monthly. Patients were followed until the sooner of 5 years or a change of therapy due to either intolerance or resistance. The primary end point of the trial was event-free survival at 5 years. The main clinical findings of the study have been presented (O'Brien et al. 2018). All patients had signed an informed consent prior to biobanking.

The samples in the CML study cohort2 were the first 141 biobanked samples in the trial plus 18 additional patients whose disease progressed. Eighty-one patients received imatinib and 78 dasatinib as their first-line treatment. Their median age was 53 years (Q_1 = 43, Q_3 = 63), median follow-up time was 60 months (range 1–60 months) and median overall survival was 60 months (95% CI, 60 to 60).

4.7.3 Other patient materials used in this thesis

4.7.3.1 Testicular cancer study material (I)

Testicular cancer patient material comprised of 20 patients' samples from the pathology archives at Turku University Hospital. Tissue samples had been collected from either seminomas or embryonal carcinomas at the time of surgery and these samples were further paraffin embedded and formalin fixed, and then used for immunohistochemistry in this study. The usage of human tissue samples was approved by the Finnish national authority for medicolegal affairs (Dnro 889/04/047/08) and regional ethics committee of University of Turku (Dnro 146/2007).

4.7.3.2 HNSCC study material (I)

HNSCC patient material included tissue samples from 52 patients from the pathology archives at Turku University Hospital. The basic clinicopathological

characteristics for these patients are summarized in Table 8 as they were not included in the original publication I. Same patient material was also used in publication (Routila et al. 2016). The median age of these patients at diagnosis was 63 years and the median follow-up times was 37 months. The most common site of the primary tumor was the oral cavity (19 patients (37%)), followed by laryngeal tumors (12 patients (23%)). Thirty-one patients (60% of the patients) did not have nodal metastasis (N class N0), but the 5-year recurrence of the disease was observed in 28 patients (54% of the patients). Twenty-nine patients (56% of the patients) had been treated with radiotherapy. Tumor samples had been collected at the time of surgery and these samples were further used for immunohistochemistry in this study. The usage of human tissue samples was approved by the Finnish national authority for medicolegal affairs (Dnro 889/04/047/08) and regional ethics committee of University of Turku (Dnro 146/2007).

Table 8. The basic clinicopathological characteristics of the 52 HNSCC patients in the HNSCC study material used in original publication I.

Characteristic	No.	%	Characteristic	No.	%
Age (years)			**N class**		
< 60	22	42	N0	31	60
≤ 60 years	27	52	N1	6	12
N/A	3	6	N2	11	21
Sex			N/A	4	8
Male	39	75	**Radiotherapy**		
Female	13	25	Yes	29	56
Smoking			No	23	44
Yes	21	40	**5-year recurrence**		
No	26	50	Yes	28	54
N/A	5	10	No	15	29
Alcohol consumption			N/A	9	17
Yes	5	10	**Survival (2015/10)**		
No	42	81	Alive	30	58
N/A	5	10	Dead	22	42
T class			**Primary site**		
T1	11	21	Oral cavity	19	37
T2	13	25	Larynx	12	23
T3	15	29	Oropharynx	6	12
T4	9	17	Tongue	5	10
			Lower lip	3	6
			Maxilla	3	6
			Other	4	8

4.7.4 End point definitions (I–III)

Overall survival (OS) was defined for all patients measured from the date of diagnosis to the date of death from any cause. Patients not known to have died at the last follow-up were censored on the date they were last known to be alive. “Time to relapse” was defined for patients from the date of diagnosis until the date of relapse. Patients not known to have relapsed were censored on the date they were last examined. Event free survival (EFS) was defined for CML patients from the date of diagnosis to the first occurrence of any of the following: death from any cause during treatment, progression to the accelerated phase or blast crisis of CML, or loss of a cytogenetic response. “Time to complete molecular response (CMR)” was defined from the date of diagnosis to the date of no detectable BCR-ABL1 transcripts in two consecutive samples with good quality control values (BCR-ABL1/ABL1 ratio of ≤ 0.0032%, in the presence of at least 31623 control ABL1 transcripts). Freedom from progression (FFP) was defined from the date of chronic phase to the date of accelerated phase or blast crisis of CML.

4.8 Statistical analysis (I–III)

Statistical analysis and comparisons were performed using SAS software (version 9.3, SAS Institute Inc., Cary, NC, USA), JMP pro (version 12.0, SAS Institute Inc.) or GraphPad Prism (version 8.3., GraphPad Software, La Jolla, CA, USA). The sample size for studies was chosen according to previous studies in the same area of research. All statistical tests were two-sided and declared significant at a p-value of <0.05.

Continuous variables were summarized by descriptive statistics (median, interquartile range and range) while frequencies and percentages were calculated for categorical data. Patients were stratified according to gene expression into high (>median expression of the studied gene in AML patients) and low (<median expression of the studied gene in AML patients). Additional analysis was performed by using overexpression (>mean expression of the studied gene in normal sample), underexpression (<mean expression of the studied gene in normal sample) or subpopulation analysis based on the distribution profile of the studied gene expression (also including quartiles). For continuous variables, normal distribution of the data was tested, and if needed and possible, transformations (log, ln, sqrt) were performed to achieve a normal distribution assumption. Kruskal-Wallis test, Mann-Whitney u-test, Wilcoxon rank sum test, one sample t-test, paired t-test and Student’s t-test were used for analyzing continuous variables. For categorical variables, frequency tables were analyzed using Fisher's exact test.

A Pearson’s pairwise correlation analysis was performed in a gene-to-gene manner and further hierarchical clustering (average linkage) was performed.

Separate logistic regression model was fit for ARPP19 and EVI1 alone and ARPP19+EVI1 together. Discriminative power of the three models was evaluated using Receiver Operating Characteristic (ROC) curves. A chi-squared test was used for comparison of AUC-values.

Univariable survival analysis was based on the Kaplan–Meier method where stratum-specific outcomes were compared using log-rank statistics. To adjust for the explanatory variables (diagnosis age, risk group stratification, FLT3-ITD status, NPM1 mutation status, expression levels of NOCIVA, CIP2A, SET, EVI1, WT1, ARPP19, TIPRL and PME-1), a Cox proportional hazards regression model was used for univariable and multivariable analysis. Type 1 approach was used whereby the additive effect of the marker was tested. In multivariable analysis, covariates were entered in a stepwise backward manner.

5 Results

5.1 CIP2A is an OCT4 target gene involved in HNSCC oncogenicity and radioresistance (I)

5.1.1 OCT4 and CIP2A contribute to radiation resistance in HNSCC (I)

An earlier study suggested a connection between stem cell regulator OCT4 and oncoprotein CIP2A in normal testicular stem cells (spermatogonia) (Ventelä, Côme, et al. 2012). CIP2A was shown to be expressed in the testicular progenitor population together with OCT4, and to promote self-renewal of the spermatogonia (Ventelä, Côme, et al. 2012). This study indicated that CIP2A expression is connected to the expression of OCT4 in progenitor cells, but how and if these genes regulate each other's expression and function remained unexplored. Furthermore, it has been demonstrated that spermatogonia cells resemble embryonic stem cells in many ways and also possess high pluripotent capacity (Kanatsu-Shinohara et al. 2004; Ko et al. 2009).

To investigate whether CIP2A and OCT4 are expressed in the radioresistant spermatogonial stem cells, a local *in vivo* irradiation experiment with mice testes was conducted (Ventelä, Mäkelä, et al. 2012). Genes whose expression was not affected by irradiation were considered as genes that are linked to radioresistance and which are mainly expressed in the radioresistant spermatogonial stem cells. We observed that irradiation substantially increased the steady state expression of *CD9* and *Plzf* genes at 96 and 144 hours after irradiation (I, Fig. 1C), which would indicate an increased proliferation and repopulation of the spermatogonia after the irradiation insult. On the contrary, the expression of the markers of more differentiated spermatogonia, *stra8* and *c-Kit*, collapsed in response to irradiation as expected (I, Fig. 1C). Interestingly, both *Cip2a* and *Oct4* expression levels were relatively stable throughout the 144-hour follow up time (I, Fig. 1C) and closely mimicked each other's expression profiles. CIP2A protein expression in the irradiated testes was also evaluated by immunohistochemical staining and also here, CIP2A protein levels remained unchanged for the 144-hour observation period (I, Fig. S1). These data

suggest that the expression of both OCT4 and CIP2A is connected to *in vivo* radiation resistance.

Next, to assess the clinical importance of these findings, we investigated the role of CIP2A and OCT4 to act as potential biomarkers for radiosensitivity in head and neck squamous cell carcinoma (HNSCC). Patients with HNSCC are widely treated with radiotherapy, but biomarkers that would predict for tumor radiation resistance are still lacking. First, we compared *CIP2A* and *OCT4* mRNA expression levels in 15 patient-derived HNSCC (UT-SCC) cell lines (I, Fig. 4B and 6A, B), with the intrinsic radiation resistance of these cell lines. The intrinsic radiation resistance information was retrieved from earlier publications with these same cell lines (Pekkola-Heino et al. 1998; Farnebo et al. 2011). When six UT-SCC cell lines with the lowest and highest *CIP2A/OCT4* expression indexes were compared to the area under the survival curve values (AUC), a trend between high *CIP2A/OCT4* expression and the intrinsic radiation resistance was detected (I, Table 3). To further validate these results, we next analyzed HNSCC patient samples for OCT4 and CIP2A expression in relation to clinical radiation resistance. In our HNSCC study material, 29 of the 52 patients (56%) had been treated with radiotherapy. According to immunohistochemical staining, 55% (16/29) of these patients had tumors that were negative for OCT4, whereas 45% (13/29) of the patients had tumors that were OCT4 positive but all of which were simultaneously positive for CIP2A. Interestingly, in the radiotherapy treated patients, OCT4/CIP2A double positivity associated with inferior 5-year overall survival (OS) when compared to OCT4 negative patient group (I, Fig. 5D). Taken together, these data propose that high CIP2A and OCT4 expression in HNSCC cells confer HNSCC tumor radioresistance.

5.1.2 OCT4 regulates CIP2A expression (I)

Testicular cancer (TC) and embryonic stem (ES) cells are widely accepted models to study mechanisms related to cellular stemness (Clark 2007). Interestingly, we observed that in immunohistochemically stained TC patient samples, CIP2A and OCT4 were co-expressed in 95% of the studied samples (19/20) (I, Fig. 3 and Table 1). To explore the possible regulatory relationship between OCT4 and CIP2A, we selected two TC cell lines derived from either embryonal carcinoma (Tera1) or seminoma (Tcam2), and murine ES cells for *in vitro* studies. We observed that the depletion of CIP2A in Tera1 and Tcam2 cells via siRNA transfections resulted in robust downregulation of CIP2A protein expression but did not affect OCT4 protein nor mRNA levels (I, Fig. 2A, E). Comparable results were obtained when mouse ES cells derived from CIP2A hypomorphic blastocyst were examined. Although *Cip2a* expression levels were below the detection level in *Cip2a* HOZ cells, *Oct4* levels

were not substantially downregulated in these cells when compared to WT cells (I, Fig. S2). These results indicate that CIP2A is not an upstream regulator of OCT4.

Next, we explored whether OCT4 instead regulates CIP2A. Interestingly, we observed that the silencing of OCT4 in TC cell lines resulted in downregulation of both protein and mRNA levels of CIP2A (I, Fig. 2B, C, E). The inhibition of CIP2A protein expression by OCT4 siRNA was even further distinct after five days (I, Fig. 2D), suggestive of a functional relevance. To explore whether OCT4 suppression and the subsequent CIP2A inhibition would result in a functional outcome, we studied phosphorylation of MYC at serine 62 (pS62MYC) levels upon OCT4 RNAi silencing. Increase in the level of p62MYC is a well-established functional outcome of CIP2A mediated PP2A inhibition in cancer cells (Khanna, Pimanda, and Westermarck 2013). Indeed, OCT4 suppression led to robust inhibition of both CIP2A and pS62MYC expression (I, Fig. 2K) and as expected, no effect on *MYC* gene expression was detected. Lastly, to investigate whether the regulatory link between CIP2A and OCT4 exists also in ES cells, we conducted experiments with ZHBTc4 murine ES model where OCT4 can be conditionally downregulated by doxycycline treatment. As already seen in TC cell lines, also in mES cells, OCT4 protein suppression was accompanied by decreased CIP2A protein expression (I, Fig. 2J).

Taken together, these data indicate that OCT4 regulates CIP2A expression at both mRNA and protein levels in TC and ES cells.

5.1.3 CIP2A is a novel OCT4 target gene (I)

To address the ability of OCT4 to function as a transcription factor and thus to regulate CIP2A expression at the promoter level, we used Promega's Dual-Glo luciferase Assay system with previously described -1802 bp CIP2A promoter fragment (Khanna et al. 2011). Tcam2 cells were transiently transfected with CIP2A promoter and luciferase constructs and the relative promoter activity was analyzed with or without OCT4 RNAi depletion. Indeed, we observed that upon OCT4 depletion, CIP2A promoter activity significantly decreased (I, Fig. 2F). *In silico* sequence analysis of the -1802 bp promoter fragment further revealed possible octamer binding elements at region -1650 to -1600 in the promoter (I, Fig. 2G red box, and Fig. S3). Thus, we next conducted experiments with a shorter -865 bp CIP2A promoter fragment from which these predicted OCT4 binding sites were absent (I, Fig. 2G). Significantly decreased promoter activity was detected with the -865 bp fragment as compared to the -1802 bp fragment (I, Fig. 2H). Next, to study the direct role of the predicted OCT4 binding sites within the -1802 fragment, a modified version of the -1802CIP2ALuc that lacked the region -1650 to -1600 (referred as -1802ΔCIP2ALuc), was cloned (I, Fig. 2G). As expected, deletion of the

putative OCT4 binding sites resulted in significant reduction of the CIP2A promoter activity (I, Fig. 2I). However, as the decrease in CIP2A promoter activity upon 1802ΔCIP2ALuc usage was smaller than the decrease detected upon OCT4 RNAi (I, Fig. 2I red line), it is plausible that other OCT4 binding sites in addition to the region mutated here exists at CIP2A promoter. Nevertheless, these data suggest that CIP2A is a novel OCT4 target gene in TC cells.

5.1.4 CIP2A and OCT4 are co-expressed in HNSCC (I)

To investigate whether the link between CIP2A and OCT4 exists in malignancies other than TC, we next investigated HNSCC. HNSCC was chosen due to the previous identification of CIP2A as a HNSCC oncoprotein (Junttila et al. 2007; Böckelman, Hagström, et al. 2011), and due to the suggested importance of OCT4 in the therapy resistance of HNSCCs (Shen et al. 2014; Albers et al. 2012; Tsai et al. 2011). We started by assessing the mRNA expression of *CIP2A* and *OCT4* in 15 patient-derived HNSCC cell lines. Interestingly, by using a linear regression analysis, we observed a significant positive correlation between *CIP2A* and *OCT4* gene expression in these cell lines (I, Fig. 4B). Next, we analyzed the protein expression of CIP2A and OCT4 in four HNSCC cell lines and observed that all of them expressed CIP2A and OCT4 at the protein level (I, Fig. 4A). However, in HNSCC cell lines, OCT4 protein levels were notably lower than in TC cell lines (I, Fig. 4A).

To better understand in what type of cell population CIP2A and OCT4 are co-expressed in HNSCC, fluorescence-activated cell sorting (FACS) experiments for three patient-derived HNSCC cell lines were performed. Previous studies had indicated that CD24+/CD44+ double positivity can be used to define a HNSCC cell population with stem cell like characteristics (Han et al. 2014). Interestingly, although all the studied cell lines were 100% positive for CD44, by using CD24+/CD44+ double positivity we were able to extract a subpopulation from all the cell lines. The number of cells extracted with this system depended on the cell line as the frequency for the subpopulation varied from 11% to 70% between the cell lines (I, Fig. 4D, E, F). Interestingly, the CD44+/CD24+ cell population displayed clearly higher OCT4 protein expression than the CD44+ bulk of the cells (I, Fig. 4G, H, I), which could be indicative of increased stemness of the CD44+/CD24+ cell. Although CIP2A was not clearly enriched in any one subpopulation, OCT4 and CIP2A were found to be co-expressed in the CD44+/CD24+ subpopulation in all the studied HNSCC cell lines (I, Fig. 4G, H, I).

Collectively, these results demonstrate that CIP2A and OCT4 are co-expressed on mRNA and protein level in HNSCC cells.

5.1.5 OCT4 positivity is linked to increased stemness of HNSCC tumors whereas CIP2A confers poor HNSCC patient survival (I)

To explore the clinical relevance of OCT4 and CIP2A co-expression in HNSCC beyond their role in radiation resistance, we next analyzed OCT4 and CIP2A expression in HNSCC patient samples in relation to overall survival and tumorigenicity (I, Fig. 5A). According to immunohistochemistry, 83% (43/52) of the HNSCC samples were CIP2A positive, whereas 37% (19/52) were OCT4 positive (I, Table 2). Interestingly, all OCT4 positive HNSCC tumors were also positive for CIP2A. As expected, also in this patient cohort, high CIP2A staining intensity conferred inferior 5-year OS when compared to low or negative CIP2A staining intensities (I, Fig. 5B). However, when OS of OCT4 positive cancers, all of which were also positive for CIP2A, was compared to OCT4 negative cancers, no significant difference between OS rates was detected (I, Supplementary Figure 4). Next, we analyzed the differentiation status of the HNSCC tumors and observed that majority, 16/19 (84%), of the CIP2A/OCT4 double positive tumors were poorly differentiated whereas only 3/19 (16%) were well-differentiated (I, Fig. 5C).

To assess whether the CIP2A and OCT4 co-expression in HNSCC cells associate with HNSCC cells' tumorigenicity, we conducted a subcutaneous xenograft experiment with HNSCC cell lines that expressed *CIP2A* either at low (UT-SCC-50) or high (UT-SCC-14) level (I, Fig. 6A). Regarding *OCT4*, UT-SCC-14 had higher expression than already known CIP2A-dependent, tumorigenic cell lines UT-SCC-9 and UT-SCC-7 (Junttila et al. 2007), whereas expression of *OCT4* in UT-SCC-50 was similar to that of UT-SCC-7 cells (I, Fig. 6B). During the five weeks observation period, all mice injected with UT-SCC-14 cells (3/3) formed large palpable tumors, whereas only 1/3 mice injected with UT-SCC-50 formed a tiny tumor (I, Fig. 6C).

Collectively, these data indicate that CIP2A alone predicts for poor patient survival in HNSCC and confers HNSCC tumorigenicity, whereas OCT4 positivity in HNSCC is associated with low differentiation grade of the tumor.

5.2 ARPP19 promotes MYC and CIP2A expression and associates with patient relapse in acute myeloid leukemia (II)

5.2.1 PP2A inhibitor protein (PAIP) mRNA expression in AML patients (II)

PP2A inhibition has been proposed as one of the AML driver mechanisms (Arriazu, Pippa, and Odero 2016), and as a possible AML therapy target (Perrotti and Neviani

2013; Meeusen and Janssens 2018). Due to this, we wanted to examine which PAIPs are relevant for PP2A inhibition in AML. First, we conducted a systematic analysis of the expression patterns of PAIPs in AML. For this, we measured mRNA expression levels of *CIP2A, PME-1, TIPRL, SET* and *ARPP19* by real-time quantitative PCR (RQ-PCR) from 80 diagnosis phase AML patients' bone marrow (BM) samples (AML study cohort1). In addition to PAIPs, gene expression levels of EVI1 and WT1, two rather well-established AML biomarkers, were analyzed in parallel. To be able to determine the degree of overexpression in AML for any gene of interest, a pooled normal BM control sample from 56 males and females was used for normalization. We observed that *SET* was overexpressed in 30% of the patients (II, Fig. 1d), *TIPRL* in 30% of the patients (II, Fig. 1e), *ARPP19* in 21% of the patients (II, Fig. 1f), *PME-1* in 4% of the patients (II, Fig. 1h) and *CIP2A* in 4% of the patients (II, Fig. 1g).

Next, we analyzed the expression redundancies and mutual dependencies of the PAIPs in the AML study cohort1 by Pearson's correlation analysis. We found that *PME-1* expression correlated with *CIP2A*, *SET* and *ARPP19* expression (II, Fig. 1i), and SET expression with *TIPRL* and strongly with *ARPP19* expression in this patient material. *ARPP19* expression additionally correlated with *WT1* and *TIPRL* gene expression in this cohort. Hierarchical clustering of the correlation matrix further indicated that *ARPP19*, *PME-1* and *SET* expression form a cluster with similar expression profiles across the AML study cohort1 (II, Fig. 1i).

5.2.2 ARPP19 is a novel oncogene in AML (II)

Overexpression of ARPP19 has been reported to play a role in tumor progression in hepatocellular carcinoma (Song et al. 2014) and glioma (Jiang et al. 2016), but ARPP19's role in AML had not been studied. Based on the results above, *ARPP19* is overexpressed in AML and it associates with *SET*. To validate ARPP19 overexpression in AML, we analyzed an independent patient cohort with 48 AML patients (AML study cohort2) and also in this cohort, *ARPP19* was overexpressed in 58% of the patients (II, Fig. S3a).

Next, to investigate ARPP19's functional role in AML, we chose four AML cell lines based on their diverse genetic backgrounds (Kasumi-1, HL-60, KG-1, MOLM-14) for functional experiments. ARPP19 and CIP2A protein expression levels varied between the cell lines, but interestingly a positive correlation between CIP2A and ARPP19 protein expression was observed in these cell lines (II, Fig. 2a). To investigate ARPP19's role in AML cell survival, ARPP19 was stably depleted by lentiviral shRNAs in KG-1 and HL-60 cell lines in which endogenous ARPP19 protein expression was high. However, indicative of the cardinal role of ARPP19 in AML cell survival or proliferation, maintenances of a long-term ARPP19 depletion

turned out to be challenging. Nevertheless, by using early cell clones that exhibited partial ARPP19 protein knockdown, we managed to record a decrease in cell viability in both ARPP19 shRNA transduced KG-1 (II, Fig. 2c, d) and HL-60 cells (II, Fig. 2f, g). Furthermore, ARPP19 depletion resulted in reduction in the proportion of KG-1 cells in M/G_2 cell cycle state (II, Fig. 2e).

To further study ARPP19's possible oncogenic role in AML, we transiently downregulated ARPP19 with siRNAs in KG-1 and HL-60 cell lines. To our surprise, acute suppression of ARPP19 in both of the cell lines resulted in downregulation of CIP2A protein expression (II, Fig. 3a, b). In addition, decreased expression of MYC and CDK1 were observed in both cell lines upon ARPP19 depletion (II, Fig. 3a, b). We were able to validate the downregulation of MYC and CIP2A protein expression upon ARPP19 depletion in the stably transduced HL-60 cells (II, Fig. S3b). Surprisingly, upon CIP2A RNAi, we also observed inhibition of ARPP19 protein levels by about 40% in both of the studied cell lines (II, Fig. 3a, b).

Taken together, these data support ARPP19's oncogenic role in AML. These data also propose a novel hierarchical co-regulation of CIP2A and ARPP19 at the protein level.

5.2.3 *ARPP19* as a novel prognostic, relapse predicting biomarker in AML (II)

To investigate the plausible clinical relevance of ARPP19 in AML, we next examined *ARPP19* expression together with the other mRNA markers involved in this study in relation to the clinical variables of the patients in AML study cohort1. In this study cohort, 68 patients achieved CR. Patients that subsequently relapsed within the follow-up time were more likely classified as ELN-2010 adverse risk group patients than patients who did not experience disease relapse during the follow-up time. We did not detect statistically significant association between the ELN-2010 genetic risk categories and PP2A inhibitor genes, including overexpressed *ARPP19* (II, Table S7). However, supportive of the oncogenic role for ARPP19 in AML, *ARPP19* expression at diagnosis was significantly lower in patients whose disease did not relapse during the follow-up than in those whose did (II, Fig. 4a). Interestingly though, no significant difference was observed in the rate of CR, treatment resistance or death during induction therapy between the patients with *ARPP19* overexpression or underexpression. This would propose that low *ARPP19* at diagnosis is more likely linked to low disease recurrence after achieved remission than to superior response to initial induction therapy. In addition to *ARPP19*, *EVI1* was the only other studied marker whose high expression at diagnosis positively associated with relapse tendency. Notably, no significant difference between the relapsing and non-relapsing patient groups in any other clinical variables

including patient's age, secondary AML, normal karyotype, alloHSCT, extramedullary disease, NPM1 mutation and FLT3-ITD gene fusion was identified in this cohort.

To analyze the association of the expression markers included in this study with "time to relapse", Kaplan-Meier estimates and Cox proportional hazards regression models for univariable and multivariable analysis were utilized. As expected, the ELN-2010 risk groups were strong indicators of shorter time to relapse in AML study cohort1, adverse risk group having the shortest time. *ARPP19, SET* and *EVI1* gene expressions were identified as the only markers significantly associated with time to relapse in this patient material. High *EVI1* expression at diagnosis was a potent indicator of shorter time to relapse (II, Fig. S4a). Interestingly, lowest quartile (Q_1) *ARPP19* expression was indicative of longer relapse free time when compared to patients with higher *ARPP19* expression (II, Fig. 4b). The 5-year relapse rate was 7% for patients in Q_1 expression of *ARPP19*, while for patients with *ARPP19* expression higher than Q_1 it was 33%. Interestingly, patients in the Q_1 *ARPP19* expression group represented all ELN-2010 risk categories (II, Fig. 4c). This further underscores the risk group independency of *ARPP19* in relapse prediction.

When Cox hazard models for univariable analysis were examined (II, Table 1), high *ARPP19*, *EVI1* and *SET* expressions at AML diagnosis were revealed as significant predictors of shorter time to relapse in AML study cohort1. In addition, multivariable Cox hazard model for relapse that included all the expression markers in this study plus age, *FLT3*-ITD status and *NPM1* mutation status at diagnosis, revealed that patient's age, *NPM1* mutation positivity, and *EVI1*, *SET* and *ARPP19* expressions were independent prognostic factors for the time to relapse (II, Table 2). Interestingly, Cox type1 analysis additionally indicated that *ARPP19* expression at diagnosis can give additional information in AML patients relapse prediction when patient's ELN-2010 risk category and *EVI1* expression are already depicted as prerequisite factors in explaining the probability of relapse. Lastly, receiver operating characteristic (ROC) analysis demonstrated that by assessing *ARPP19* expression together with *EVI1* expression at AML diagnosis, more accurately predictions for relapse probability can be produced than by *EVI* evaluation alone (II, Fig. S4b).

Collectively, these data identify *ARPP19* expression as a potential AML biomarker that is independent of the ELN-2010 risk classification and associates with relapse tendency. Our data also indicate that evaluation of *ARPP19* at AML diagnosis could provide additional relapse predictive value to the already used diagnostic approaches.

5.2.4 *ARPP19* as a novel prognostic biomarker in AML (II)

To study *ARPP19*'s role as a prognostic biomarker in AML, we used Cox proportional multivariable hazard model for investigating the association between OS and the clinical and expression markers in the AML study cohort1. The model included age, *FLT3*-ITD status and NPM1 mutation status at diagnosis, and the expression levels of *ARPP19, CIP2A, SET, TIPRL, PME-1, EVI1* and *WT1* at diagnosis. After excluding the non-significant markers (II, Table 2), patient's age and *NPM1* mutation positivity at diagnosis as well as *EVI1* and *ARPP19* expression were identified as independent prognostic factors for OS in this patient material. Notably, *ARPP19* expression was found as the only PAIP to act as an independent prognostic factor for OS, with a hazard ratio (HR) even higher than either *EVI1*'s or age at diagnosis. To validate these results in an independent AML patient cohort, we examined the correlation between *ARPP19* expression and OS in the TCGA LAML dataset by using UCSC Xena Browser. The median served as a cut-off value for *ARPP19* high and *ARPP19* low groups. In line with our previous results, *ARPP19* expression at diagnosis was identified as a prognostic marker for OS in the LAML patient material (II, Fig. 4e). Patients with high *ARPP19* expression had significantly inferior OS as compared to patients with low *ARPP19* expression.

Collectively these results highlight a novel role for *ARPP19* as a prognostic biomarker in AML.

5.2.5 *ARPP19* expression correlates with AML disease activity (II)

Lastly, we wanted to investigate whether *ARPP19* expression levels change during the different disease phases of AML, including diagnosis, remission and relapse. If so, *ARPP19* expression could possibly be used, in addition to the already identified predictive and prognostic role, as a monitoring biomarker in AML. For this purpose, we analyzed samples from nine patients from the AML study cohort2 for which, in addition to samples at diagnosis, samples at first remission and/or at relapse were available (II, Fig. 4f). Three patients had a complete set of diagnosis, remission and relapse samples (II, Fig. 4g). Interestingly, at diagnosis, *ARPP19* was overexpressed in seven out of nine patients, but at remission, *ARPP19* expression dropped below the normal BM expression level in all the samples (II, Fig. 4f). At relapse, *ARPP19* was found to be overexpressed again in four out of eight samples. The fluctuation of *ARPP19* expression was even more evident in the complete matched set of samples, where *ARPP19* was overexpressed in all of the patients at diagnosis and relapse, and its' expression dived below the normal BM level in all of the patients at remission (II, Fig. 4g). These results thus indicate that *ARPP19* expression associates with AML disease activity.

5.3 Discovery of NOvel CIP2A VAriant (NOCIVA) and its clinical association with myeloid leukemias (III)

5.3.1 Discovery of Novel CIP2A Variant (NOCIVA) mRNA (III)

CIP2A (gene alias *KIAA1524*) is a potent human oncoprotein with extensively documented clinical relevance in several human cancers. Surprisingly, nothing is known at the moment about CIP2A mRNA and/or protein variants, not even if there is any. Thus, in this study we wanted to investigate the potential variant forms of CIP2A and started the project by employing rapid amplification of cDNA ends PCR assays (3'RACE and 5'RACE) with human cell line mRNA samples (PNT2, MDA-MB-321, HeLa). The full length CIP2A mRNA contains 21 exons and in line with our database searches for CIP2A mRNA variants, one identified mRNA variant only contained exons 1 to 19 from the CIP2A gene. In addition, a novel CIP2A mRNA splice variant, designated as NOCIVA, with an alternative exon inclusion was discovered (III, Fig. 1A). NOCIVA, therefore, became the main focus of this project. What makes NOCIVA especially interesting is that it comprises of exons 1 to 13 of normal CIP2A mRNA fused at the 3'-end to a part of an intron between exons 13 and 14 in the *KIAA1524* gene (III, Fig 1A). This 349 nucleotide long intronic region (III, Fig 1A, Fig S2B) is flanked by GT and AG nucleotides which makes it a GU-AG intron (III, Fig. S2B yellow) and which is also indicative of the AS origin of the *NOCIVA* mRNA. Interestingly, *NOCIVA* mRNA also contains a stop codon and a 330 nucleotide long 3´UTR (III, Fig. 1B and Fig. S2A).

To validate the existence and expression of a full length *NOCIVA* mRNA, PCR runs were executed in the HeLa cell line with forward primer targeting CIP2A exon1 5'-end together with reverse primers targeting the *NOCIVA* specific 3'-end of the mRNA (III, Fig 1D and Fig. S3A for PCR assay). PCR products were run on agarose gel, the appropriate size bands extracted, and the extracts subsequently sequenced to confirm the presence of the desired *NOCIVA* mRNA sequence. In addition, expression of *NOCIVA* was confirmed in several cancer cell lines with PCR primers specific to the novel coding sequence of NOCIVA (III, Fig. 1E and Fig. S3B for PCR assay).

In silico analysis with Human Splicing Finder (version 3.1, (Desmet et al. 2009)) identified multiple exonic splicing silencer (ESS) matrices, particularly PESS-octamers and Fas ESS, at the junction site between NOCIVA and CIP2A. Furthermore, SpliceAid 2 (Piva et al. 2012) and SFmap (version 1.8, (Paz et al. 2010)) identically predicted binding of SRp20 (SRSF3) and YB-1 splicing factors at the *NOCIVA* junction site (III, Fig. 1C), both of which have been reported to foster

exon-inclusion during AS (Wei et al. 2012; Xiao et al. 2016). SpliceAid 2 and SFmap reported also multiple additional splice factor binding sites in the near vicinity of the junction site (III, Fig. 1C). These splice factors included SRp40, SLM2, Sam68 and multiple hnRNPs, including hnRNP K.

5.3.2 Characterization of NOCIVA protein (III)

The potential NOCIVA protein contains 545 amino acids similar to those in CIP2A, followed by the NOCIVA specific peptide sequence that contains 13 amino acid long stretch NNKNTQEAFQVTS at the C-terminus (III, Fig. 1B). Altogether in NOCIVA protein there would be 558 amino acids, whereas full length CIP2A protein contains 905 amino acids. Interestingly, when we conducted a Blast homology search (Altschul et al. 1997) (III, Fig. S4A, BLASTP 2.8.1+, Database: Non-redundant protein sequences (nr)) for the 13 amino acid long peptide sequence specific for NOCIVA, no match was found within the human proteome. This enabled us to produce NOCIVA specific antibodies. First, the specificity of the two anti-NOCIVA antibodies was examined with the help of recombinant NOCIVA and CIP2A proteins. As shown in original publication III figure 2A, anti-NOCIVA antibodies exclusively recognize NOCIVA but not the full length CIP2A protein nor any of the CIP2A protein fragments. The NOCIVA specific signal could also be abolished by using blocking peptide (III, Fig. S4B for NOCIVA ab #2 data). Next, to examine the spatial expression of the endogenous NOCIVA, we executed immunofluorescence (IF)-staining in MDA-MB-231 cells. As expected, CIP2A was found to reside primarily in the cytoplasm, whereas surprisingly signal for endogenous NOCIVA was mainly nuclear (III, Fig 2B and Fig. S4C). Also, when NOCIVA was overexpressed as a GFP fusion protein, NOCIVA-GFP colocalized with DAPI to the nucleus (III, Fig. 2C). Collectively these data indicate that NOCIVA contains a novel and immunogenic peptide sequence, and accounts for a nuclear CIP2A variant protein.

Next, to study NOCIVA protein functions, we compared recombinant GST-CIP2A 1-560 and GST-NOCIVA fusion proteins (III, Fig. S4D for Coomassie staining) in two central functions for CIP2A mediated PP2A regulation: in direct binding to PP2A-B56α subunit, and in protein homodimerization. CIP2A 1-560 protein was chosen as we know from previous work in our laboratory that the full length CIP2A recombinant protein is highly unstable and thus hard to work with, as the structural domains needed for CIP2A functions reside within the first 560 amino acids, and as the difference between NOCIVA and CIP2A 1-560 mainly exists in the NOCIVA specific 13 amino acid at the C-terminal end. As anticipated, both of the recombinant proteins co-immunoprecipitated B56α with equivalent efficiency (III, Fig. 3A). This can be explained by the fact that the B56α binding regions locate in

the N-terminal part of CIP2A which remains similar to that of CIP2A in NOCIVA protein (Wang et al. 2017). NOCIVA was also capable to heterodimerize with CIP2A 1-560, although with weaker affinity than what was seen with CIP2A 1-560 homodimers (III, Fig. 3B). This is also logical as we know that the CIP2A-NOCIVA fusion site partly overlaps with the CIP2A protein sequence that is needed for CIP2A homodimerization (Wang et al. 2017) (III, Fig. 3C,D). Therefore, when comparing to CIP2A homodimers, in NOCIVA (III, Fig 3D,E) some of the stabilizing amino acid interactions are lost (Fig 3E; for example, hydrogen bond and salt bridges between Q559-E560 and S519-Q553-R557-D520-Y556), but also a novel stabilizing interaction network is generated (for example, hydrogen-bond network with N547-Q551-Q555). On the other hand, residues distinct between CIP2A and NOCIVA, which are actually also evolutionary highly conserved in CIP2A, create more charged dimer interface in the latter.

Taken together, these results demonstrated that NOCIVA protein is able to directly bind to the PP2A-B56α subunit and to heterodimerize with CIP2A.

5.3.3 *NOCIVA* expression in normal and cancer cells (III)

To assess the expression levels of *NOCIVA* and *CIP2A* in cell and patient samples, two quantitative real time PCR (RQ-PCR) assays for both *NOCIVA* and *CIP2A* (III, Fig. S5A for PCR assays) were designed and validated. Both *NOCIVA* RQ-PCR assays were designed to amplify the *NOCIVA* specific mRNA sequence (NOCIVA#1 and NOCIVA#2 assay). The *CIP2A* RQ-PCR assays were designed to amplify the exon13-14 (CIP2A e13 assay) or exon20-21 (CIP2A e20 assay) branch site. The specificity and amplification efficiency of the designed RQ-PCR reactions were verified by agarose gel electrophoresis, melting curve and standard curve analysis with the HeLa cell line, clinical AML samples and the pooled normal BM sample (III, Fig. S1A-J). A single peak in the melting curve analysis and a single band of the expected size on the gel, were required. The amplification efficiency of all used assays, including control genes beta-actin and GAPDH, was 90-100%, which allowed a direct comparison between the expression levels of the genes examined in this study. NOCIVA#1 and CIP2A e20 were the mainstay assays in this study when referring to *NOCIVA* and *CIP2A* expression.

First, we wanted to evaluate *NOCIVA* and *CIP2A* expression in normal human tissues and for this we used commercially available normal human tissue cDNA panels (Human MTC panel I & II, Clontech, cat no 636742 & 636743). In line with previously published data, *CIP2A* was expressed at low levels in most normal human tissues, except for testis (III, Fig. S5B). Intriguingly, CIP2A e13 assay showed markedly higher expression levels than CIP2A e20 assay in many tissues. This proposes that the full-length *CIP2A* is not the dominant *CIP2A* isoform expressed

across all human tissues and maybe other *CIP2A* mRNA variants exist. *NOCIVA* was also expressed at low levels in most normal human tissues (III, Fig. 4A), with a tissue-specific expression profile comparable to that of *CIP2A*, including the high expression in testis. Next, we wanted to identify tissues in which *CIP2A* AS to *NOCIVA* could be more profound and for this counted a ratio between *NOCIVA* and *CIP2A* expression in the normal tissues (III, Fig. 4B). Kidney, leukocytes, and pancreas were identified as the tissues with the highest *NOCIVA/CIP2A* ratio. Next, to examine the possibility of elevated expression of *NOCIVA* in cancer, we analyzed the expression of *NOCIVA* and *CIP2A* in a patient derived normal epidermal keratinocyte (NHEK) and HNSCC cell line panel. Intriguingly, in addition to expected overexpression of *CIP2A* (III, Fig. S5C), also *NOCIVA* expression was significantly increased in HNSCC as compared to NHEK (III, Fig. 4C).

Following the highest *NOCIVA/CIP2A* ratio identified in the normal leukocytes, we next examined whether this preferential *NOCIVA* expression could also be seen in myeloid cancer cells. Interestingly, relatively higher expression of *NOCIVA* over *CIP2A* was indeed detected in most of the AML and CML cell lines analyzed (III, Fig. 4D). NOCIVA protein expression was also confirmed in two AML cell lines, Kasumi-1 and KG-1, by Western blotting (III, Fig. 4E). Encouraged by these results, we next set to investigate the preferential *NOCIVA* expression from AML study cohort1 (n=80) and CML study cohort1 (n=35) patient materials. In line with previously published results (Lucas et al. 2011; Mäkelä et al. 2019), 96% of AML patients and 94% of CML patients expressed *CIP2A* at lower levels than in normal BM control (III, Fig. 4F,G). In contrast and supportive for the active AS of *CIP2A* to *NOCIVA* in myeloid cancers, *NOCIVA* was overexpressed in 77% of AML and 65% of CML patients (III, Fig. 4 F,G). In addition, to study the mutual dependencies in expression levels of *NOCIVA*, *WT1*, *EVI1*, *CIP2A, SET, ARPP19, TIPRL* and *PME1* in AML, Pearson pairwise correlation analysis was conducted. In the AML study cohort1, *NOCIVA* expression significantly correlated with *PME1* and weakly but significantly also with *ARPP19* and *SET* gene expression (III, Fig. 4H).

Taken together, these data provide evidence that *NOCIVA* has an expression pattern similar to that of *CIP2A* across normal human tissues, but in contrast to *CIP2A*, *NOCIVA* displays robust overexpression both in AML and CML at the mRNA level.

5.3.4 Clinical relevance of *NOCIVA* expression in AML (III)

Our results above propose that the AS of *CIP2A* to *NOCIVA* is exceptionally active in AML and CML. Yet the possible clinical relevance of *NOCIVA* transcript in these cancers was unclear. To address this question, we set to investigate the prognostic significance of *NOCIVA* expression in the AML study cohort1 patient material. To

examine the association of OS with *NOCIVA*, *CIP2A* and the other expression markers already involved in original publication II, Kaplan-Meier estimates and Cox proportional hazards regression models for multivariable analysis were conducted. Expression of all the studied marker was divided into high and low according to median expression of a given marker. Interestingly, based on Kaplan-Meier estimates, high *NOCIVA* expression was identified as a strong indicator of poor OS in this AML patient material (III, Fig 5A). On the contrary, low *CIP2A* expression was only a borderline significance predictor of better OS in this cohort (III, Fig 5B). Moreover, the Cox proportional multivariable hazard model, which included age at diagnosis and expression levels of *PME1, SET, TIPRL, ARPP19, EVI1, WT1, CIP2A e13, CIP2A e20* and *NOCIVA* at AML diagnosis, indicated that age at diagnosis, and *EVI1* and *NOCIVA* expression were independent prognostic factors for OS in this study material (III, Fig. 5C). Interestingly, the hazard ratio of *NOCIVA* expression (HR: 1.51) was even higher than of either *EVI1* expression (HR: 1.27) or age at diagnosis (HR: 1.07), both of which have well-establish roles as strong predictors of AML outcome. Also in this analysis, low rather than high expression of *CIP2A e13* was a borderline significance predictor of poor OS. In addition to OS, we also evaluated the association of the studied expression markers with clinical variables and ELN-2010 genetic risk categories. The expression of *NOCIVA* or *CIP2A* did not correlate with any of the following clinical variables: gender, age, secondary leukemia, the presence/absence of normal karyotype, or leukocyte or BM blast count. Furthermore, neither *CIP2A* nor *NOCIVA* expression levels associated with the ELN-2010 genetic risk categories (III, Fig. 5D).

Collectively these results identify a pronounced and ELN-2010 genetic risk category independent association between high *NOCIVA* expression and adverse AML patient outcome.

5.3.5 Clinical relevance of *NOCIVA* expression in CML (III)

Next, to assess whether *NOCIVA* mRNA expression has a prognostic role also in CML, we analyzed *NOCIVA* expression in association to OS and event free survival (EFS) in 34 newly diagnosed CP CML patients (CML study cohort1). Out of these, 20 patients had received imatinib and 14 patients either dasatinib or nilotinib as the frontline treatment. As calculation of OS was not reasonable in this study cohort due to only one death during the follow-up time, we used Kaplan-Meier estimates to analyze the EFS for these patients. The median *NOCIVA* expression was again used for stratifying patients into high or low *NOCIVA* expression groups. Interestingly, high *NOCIVA* expression was found to be associated with significantly shorter EFS (III, Fig. 6A). Moreover, EFS was even more inferior in the imatinib treated high *NOCIVA* patient group (III, Fig. 6B). Surprisingly this was exclusive for imatinib

treated patients as this was not seen in the patients who received 2G TKIs as the first-line treatment (III, Fig. 6C). In addition to EFS, we also evaluated "Time to Complete Molecular Response (CMR)" to examine the depth of patients' response. CMR was considered as the deepest form of response. Interestingly, among the imatinib treated patients, high *NOCIVA* expression significantly associated with shorter time to CMR (III, Fig 6D). To be more exact, none of the patient with high *NOCIVA* at CP CML diagnosis achieved CMR within the 80 months follow-up time. On the contrary, among the patients treated with 2G TKIs, no association was detected between time to CMR and *NOCIVA* expression.

To further validate the prognostic or predictive role of *NOCIVA* in CML, we analyzed another independent CML patient cohort (CML study cohort2) for NOCIVA expression in relation to disease progression. This patient material consisted of 159 patients from the SPIRIT2 clinical trial (O'Brien et al. 2018) and 81 patients had received imatinib and 78 dasatinib as their frontline treatment. Consistent with the results from CML study cohort1, also in this patient material high *NOCIVA* expression at CP CML diagnosis associated with disease progression exclusively among the patients treated with imatinib. Among these patients, those whose disease eventually progressed to blast crisis, had significantly higher *NOCIVA* expression at diagnosis (III, Fig. 6E). In addition, among the imatinib treated patients, highest quartile *NOCIVA* expression significantly associated with shorter freedom from progression (FFP) (III, Fig 6F). Also in line with the data from CML study cohort1, no association between FFP and *NOCIVA* expression could be detected among the patients treated with dasatinib (III, Fig. 6G).

Taken together, these results identify a clinically relevant association between high *NOCIVA* expression at CP CML diagnosis and shorter FFP, EFS and lower rates of CMR exclusively among the patients treated with imatinib as the first-line therapy.

6 Discussion

6.1 Identification of novel roles for CIP2A in HNSCC

Stemness of the cancer cells has been established as an important phenomenon due to its strong association with inferior outcome in a wide spectrum of cancers (Ng et al. 2016; Tirosh et al. 2016; Smith et al. 2018; Malta et al. 2018). One of the underlying reasons for poor cancer patient survival, that has also been closely linked to the stem cell nature of cancer cells, is the substantial resistance among patients against DNA damaging anticancer therapies, including radiotherapy. Regardless of the underlying mechanism for the observed stemness, whether it reflects the presence of CSCs or the adoption of stem cell-associated programs by the non-CSC cancer cells or maybe both, transcription factors highly expressed in embryonic stem cells, such as OCT4, have been reported to be re-expressed or reactivated in several cancers. In this thesis we identified CIP2A as a novel OCT4 target gene associated with radiation resistance and tumorigenicity of HNSCC cells. The radiation resistance phenotype of CIP2A/OCT4 double positive cells was demonstrated by multiple different approaches extending from radioresistant, normal spermatogonia stem cell population *in vivo* to both HNSCC cell lines and tumor tissues *in vitro*. Together these results may provide a novel basis for prediction of HNSCC patient response to radiotherapy. Future identification of OCT4 driven CIP2A target mechanisms may help in further understanding of radioresistance in HNSCC and in other cancers, as well as development of novel radiosensitation therapies.

Connection between OCT4 and CIP2A was initially observed in normal testicular stem cells (Ventelä, Côme, et al. 2012). In this thesis, we demonstrated with *in vivo* irradiation experiment that similar to *bona fide* stem cell marker *Oct4*, also *Cip2a* is expressed in the radiation resistant subpopulation of mouse testicular cells. In testicular cancer, we reported a co-expression of OCT4 and CIP2A in 95% of the patient samples, and OCT4 was found to unidirectionally regulate CIP2A expression at both mRNA and protein levels. To investigate whether the relationship between CIP2A and OCT4 exists in other cancers than of testicular origin, we used patient derived HNSCC cell line and patient samples. HNSCC was chosen based on previous identification of CIP2A as an oncoprotein in HNSCC (Junttila et al. 2007;

Böckelman, Hagström, et al. 2011), and suggested importance of OCT4 in the therapy resistance of HNSCCs (Shen et al. 2014; Albers et al. 2012; Tsai et al. 2011). Indeed, our results demonstrated that CIP2A and OCT4 are co-expressed in HNSCC and a positive correlation between *OCT4* and *CIP2A* mRNA expression in HNSCC was detected. Importantly, all OCT4 positive HNSCC tumor samples were also positive for CIP2A. Our experiments further demonstrated that patient derived HNSCC CD24/CD44 double positive CSC-like cells expressed both OCT4 and CIP2A, suggesting a possible clinical role for their interaction. Lastly, a trend between high *CIP2A/OCT4* expression and the intrinsic radiation resistance of the HNSCC cell lines was established. Collectively, these results provide first ever evidence for the relationship between OCT4 and CIP2A in HNSCC and the relevance of this relationship in cancer.

OCT4 is a well-established stem cell transcription factor, whose overexpression in malignant cells has been demonstrated to correlate with poor patient survival and therapy resistance in several cancer types (Shen et al. 2014; Tsai et al. 2011). In radiation resistance, the role of OCT4 and other *bona fide* stem cell factors has traditionally been linked to their ability to maintain DNA damage resistant stem cell populations from which the regenerating progenitor cells emerge after an irradiation insult. However, OCT4 target genes contributing to radioresistance via modulating apoptosis resistance and cell proliferation have been poorly understood. Identification of CIP2A as a novel OCT4 target gene thus foster a better understanding of the mechanisms by which OCT4 regulates radiation resistance. As we know that CIP2A mediated inhibition of PP2A boosts multiple oncogenic mechanisms such as Akt kinase activity and phosphorylation of both MYC and E2F1 (Khanna, Pimanda, and Westermarck 2013; Laine et al. 2013), in the future it would be interesting to study whether these CIP2A effector pathways are regulated by OCT4, and whether they have any functional relevance in OCT4 driven radiation resistance. Also, as we know that CIP2A is overexpressed in vast majority of human cancers, and its expression predicts adverse patient survival in a large number of different cancers (Khanna, Pimanda, and Westermarck 2013), it is expected that the results of this study are not unique to HNSCC but could have clinical implications in the treatment of several other cancers in addition to HNSCC. Collectively, we postulate that it is actually the combination of OCT4 being a regulator of cellular stemness and a regulator of phosphoprotein signaling via CIP2A, that confer the clinically relevant radioresistance identified in this study.

Results of this study support a conclusion that CIP2A plays an important role in the determination of radiation resistance of HNSCC cells. The well-established role of OCT4 in mediating radiation resistance together with the data from our study implies that particularly OCT4 driven expression of CIP2A is relevant for the radiosensitivity of the HNSCC tumors. In fact, in rectal cancer, a direct suppression

of CIP2A by RNAi was recently shown to result in radiosensitation of colorectal cancer cells *in vitro* (Birkman et al. 2018). These conclusions are further supported by a study where CIP2A was shown to be expressed in crypts of mouse intestinal cells and needed for effective intestinal regeneration in response to both irradiation and DNA damaging cisplatin therapy (Myant et al. 2015). Interestingly, OCT4 driven CIP2A expression in testicular cancer cells investigated in this study as well as CIP2A expression in irradiated mouse intestine cells (Myant et al. 2015), induced the expression of the serine 62 phosphorylated oncogenic form of MYC. Moreover, because checkpoint kinase CHK1 inhibition has been shown to decrease the transcription of CIP2A (Khanna et al. 2013) and recently, OCT4 linked expression of CHK1 was shown to promote radioresistance of rectal cancer cells (Shao et al. 2018), it is plausible that both CHK1 and OCT4 play a role in CIP2A mediated radiation resistance. On the other hand, as multiple studies have already linked high CIP2A expression to promotion of resistance or poor patient survival after treatment with DNA damaging anticancer drugs (Choi et al. 2011; Böckelman, Lassus, et al. 2011; Laine et al. 2013), it is clear that in addition to its' importance in radiation resistance, CIP2A has a more general role in determining the response to DNA damaging therapies. This in mind, the results from Myant et al. study suggest that pharmaceutical inhibition of CIP2A might have unfavorable effect on the regeneration of normal tissues when combined with DNA damaging anticancer agents (Myant et al. 2015). Therefore, it would be most reasonable to combine inhibitory targeting of CIP2A with localized, exclusively tumor tissue targeting, radiotherapy. Lastly, in this study we demonstrated that high CIP2A expression was linked to poor 5-year OS in HNSCC, whereas inhibition of CIP2A has already been shown to effectively inhibit HNSCC tumorigenesis (Junttila et al. 2007). These results imply that targeting of CIP2A in HNSCC, but maybe also in other cancers, could simultaneously be used for elimination of the OCT4 negative bulk of the tumor as well as for radiosensitation of the OCT4 positive CSC-like cells.

Multiple promising biomarkers have been identified in HNSCC, yet none of the evaluated biomarkers have gained enough clinical support for them to be used as a part of the routine clinical practice (Hsieh et al. 2019). This means that there are at the moment either no validated biomarkers available for the prediction of treatment response in HNSCC. In this study, we identified OCT4 and CIP2A as novel biomarkers to predict radioresistance of HNSCC. Our observation that CIP2A/OCT4 double positivity predicts for poor patient survival in HNSCC patients treated with radiotherapy is interesting as predictive biomarkers for poor radiotherapy response in HNSCC are lacking. It is however clear that further examination using larger HNSCC patient materials is needed to confirm the clinical usefulness of our results. In a meta-analysis study of 27 studies including 2143 patients by Fan et al., evidence of four stemness markers, CD133, NANOG, BMI-1 and OCT4, in patients with

HNSCC was combined to assess the prognostic value of these markers (Fan et al. 2017). Interestingly also our study was included in this analysis. From this study they concluded that high expression of these markers served as a prognostic factor for lower DFS and OS, but these markers were rather linked to relapse of HNSCC than death from HNSCC. As was done in Fan et al. study, it seems that there is an ever-increasing trend of combining markers to definitive signatures and indexes to be used in patient stratification. In many cases this is reasonable as instead of using single biomarkers, signatures or indexes can be adopted to various uses more easily (diagnostic, predictive, prognostic, monitoring, and even from one cancer to another) and give a more comprehensive picture of the disease with a single analysis. With this in mind, maybe in the future the diagnostic assessment of CIP2A in-parallel with biomarkers indicative of cancer cell stemness would be more powerful approach for HNSCC tumor radiation resistance evaluation, than either CIP2A or OCT4 alone.

6.2 ARPP19 as a novel oncogene and prognostic biomarker in AML

The diagnostic workup for AML includes screening for multiple molecular markers that assign patients to ELN genetic risk categories (Döhner et al. 2010; Döhner et al. 2017). These markers together with clinical parameters such as patient's age and performance status have been associated with treatment paths and prognosis yet fall short in accurately predicting patient outcomes (Walter et al. 2015). Prognostic markers that are used today are also poor in explaining the high prevalence of relapse among AML patients, and up to 50% of patients will relapse after first remission. Therefore, in order to improve outcome of AML patients, novel insights of the molecular mechanisms behind AML as well as risk factors predictive for the disease relapse have been awaited. A better understanding of mechanisms affecting AML relapse tendency independently of the currently used risk categories would also be of high medical relevance. In this thesis study, we surveyed associations of PP2A inhibitor protein (PAIP) gene expression at AML diagnosis with clinical outcomes of the patients. Our results identified ARPP19 as a novel mRNA biomarker for estimation of low relapse risk in patients with AML after standard induction therapy. In addition, ARPP19 was identified as a novel oncogenic protein in AML. ARPP19 depletion in AML cell lines decreased cell viability and inhibited expression of oncoproteins MYC, CDK1 and CIP2A. Importantly, both *ARPP19* expression and its association with relapse risk in AML patients was independent of the ELN-2010 risk group classification. The relapse predictive role of *ARPP19* was also additive when patient's *EVI1* expression and genetic risk group were taken into account. Our results thus indicate that measurement of *ARPP19* expression at AML diagnosis could yield clinically relevant additional value to the currently used diagnostic

approaches. Our results also suggest that *ARPP19* expression at diagnosis could maybe be used in the future for patient stratification strategies that guide patients with low relapse risk and low *ARPP19* expression to chemotherapy, whereas patients with high relapse risk and high *ARPP19* expression should be from the beginning treated more intensively, such as with HSCT.

ARPP19 is highly expressed in embryonic tissues, and its expression decreases progressively during development, suggesting for stem cell nature of ARPP19´s physiological expression (Girault et al. 1990). Interestingly, a link to LSCs also already exists, as ARPP19 has been reported to contribute to a LSC signature in two independent studies (Eppert et al. 2011; Vitali et al. 2015), but notably not to hematopoetic stem cell signature. ARPP19 was also one of the three genes involved in the phenotypic LSC signature that predicted adverse prognosis in an AML patient cohort with 110 patients (Eppert et al. 2011). However, the independent role of ARPP19 in AML has not been investigated before this study. Neither has the risk group independent role of ARPP19 in AML relapse prediction been demonstrated before this study. In this study, patients with *ARPP19* overexpression at diagnosis were more likely to be classified as FAB M1 subtype, which is AML with minimal maturation. Additionally, we observed a link between high *ARPP19* expression and minimal maturation level of the leukemic cells in TCGA LAML dataset, where highest *ARPP19* expression was associated with M0 AML subtype. I therefore hypothesize that high *ARPP19* expression at AML diagnosis might be an indication of the presence of therapy resistant LSC or other cancer stem cell like population from which the disease relapse subsequently arises in these patients. Functionally the association of ARPP19 with non-differentiated stem cell like cells is supported by findings that down-regulation of ARPP19 gene by miR-451 overexpression associates with erythroid maturation and differentiation (Bruchova-Votavova, Yoon, and Prchal 2010). On the other hand, in gastric cancer cells, ARPP19 was recently reported to upregulate expression of CD44 and to promote sphere formation, which indicated the enhancement of cancer stem cell like properties (Gao et al. 2020). Together these results propose that ARPP19 could be involved in cancer cell stemness, but naturally further research is needed to unravel the exact role of ARPP19 on this topic.

Decreased tumor suppressor activity of PP2A due to elevated expression of PAIPs has been demonstrated to promote malignant growth of several cell types (Khanna and Pimanda 2016; Kauko and Westermarck 2018), including myeloid lineage leukemia cells (Lucas et al. 2011; Arriazu, Pippa, and Odero 2016). In AML, SET has been shown to promote both malignant growth and drug resistance (Yang et al. 2012; Agarwal et al. 2014), and CIP2A inhibition has been reported to reduce proliferation and MYC expression (Barragán et al. 2015). The central role of PP2A inhibition in AML (Arriazu, Pippa, and Odero 2016; Haesen et al. 2014) and in other

cancers (Kauko and Westermarck 2018), provides a clear scientific rationale for the clinical association between low *ARPP19* expression and low relapse tendency of AML discovered in our study. The proposed oncogenic role of ARPP19 in AML is further supported by our demonstration that upon ARPP19 depletion, the expression of MYC, a well-validated oncogenic target of PP2A, decreases. Intriguingly, our data also demonstrated that ARPP19 positively regulates CIP2A protein expression in AML cells although we did not detect any significant correlation between *ARPP19* and *CIP2A* mRNA expression in our AML patient material. Therefore, it is possible that similarly to CML (Lucas et al. 2011; Lucas et al. 2015), also in AML, CIP2A is regulated at protein level. In fact, in a recent study, high CIP2A protein levels were found to correlate with poor survival from relapse in normal karyotype AML patients, whereas no correlation was detected between CIP2A mRNA levels and clinical outcome (Lucas et al. 2018). Thus, further research on the ARPP19 mediated regulation of CIP2A protein expression in AML is clearly warranted. Lastly, the functional hierarchy between ARPP19 and CIP2A proteins established in this study provides a novel view on why ARPP19 might have a stronger clinical role in AML than CIP2A. The rationale here is that in addition to control of its own direct PP2A-B55-subunit targets (Gharbi-Ayachi et al. 2010), ARPP19 is able to control PP2A-B56-subunit targets (Wang et al. 2017) via CIP2A regulation. Hence, maybe the decreased PP2A activity in AML due to ARPP19 overexpression could be restored by therapeutically blocking ARPP19's oncogenic effects on PP2A. Even so, prior to development of any ARPP19 targeted therapies, ARPP19 protein structure needs to be solved. The first step towards the structural level understanding of ARPP19 function was recently taken when the backbone assignment of the human ARPP19 protein was published (Thapa et al. 2020).

The ability to measure residual disease (MRD) below the morphology-based 5% blast threshold has been established as an important tool for refining the risk categories in AML (Schuurhuis et al. 2018). MRD assessment in AML is justified since it can among other things improve the risk assessment, provide an objective method to establish a deeper remission status, and identify emerging relapse and thus enable early intervention (Schuurhuis et al. 2018). Currently two methods are widely applied for MRD detection, RQ-PCR and multiparameter flow cytometry (MFC). Newer technologies including NGS and digital PCR are also emerging. However, PCR based molecular MRD assessment, which is of high sensitivity and hence generally considered the gold standard, applicability is limited to only 40% of AML patients that carry one or more appropriate genetic alterations (Hourigan et al. 2017). Based on our results, ARPP19 should be further validated as a potent, novel MRD biomarker in AML. Our data showed that *ARPP19* mRNA expression levels followed the disease activity in patients who achieved remission with conventional induction therapy and whose disease subsequently relapsed. In the future, it would

be especially interesting to evaluate the potential usefulness of ARPP19 as an MRD monitoring biomarker that could be followed up in patients postremission to identify the potentially arising relapse. Even though mRNA expression assessment has been earlier considered as a challenge in MRD monitoring, advancements in AML patient sample digital PCR assays (Voso et al. 2019) and especially the digital droplet PCR assays already in clinical use for monitoring disease activity of CML by analyzing *BCR-ABL1/ABL1* mRNA expression ratios (Nicolini et al. 2019) imply that mRNA expression assessment will be feasible also for patients with AML in the near future.

6.3 NOCIVA as a novel prognostic and predictive biomarker in myeloid leukemias, AML and CML

CIP2A has clinical relevance in a great number of human cancers (Khanna, Pimanda, and Westermarck 2013), and is an attractive cancer therapy target due to its direct inhibition of PP2A-B56α (Wang et al. 2017) and low expression in normal human tissues (Junttila et al. 2007). Surprisingly, apart from a handful of splice variant and protein variant predictions listed in databases such as NCBI The Nucleotide and The Protein database, nothing is currently known about mRNA or protein variants of CIP2A. In my thesis study I identified a novel CIP2A splice variant, NOCIVA, which possesses properties of a clinically relevant PP2A inhibitor in myeloid malignancies. In AML, NOCIVA could potentially in the future serve as a prognostic biomarker that provides clinically relevant additional predictive value for patient outcomes, whereas in CML, it could be used as a predictive biomarker that guides TKI treatment decisions between imatinib and 2G TKIs. Importantly, none of the predicted CIP2A splice variants in the databases resemble NOCIVA, but neither have any of the predicted variants been examined for their functional or clinical relevance. This notion both underlines the absolute novelty of my study as well as points out the obvious need for experimental validation of these isoforms in order to gain a comprehensive understanding of the regulation and function of CIP2A in cancer. Further studies to validate the diagnostic value of *NOCIVA* mRNA as a novel biomarker in myeloid leukemias are clearly warranted.

NOCIVA protein is generated when a NOCIVA specific 13 aa long peptide tail is added to the CIP2A aa number 545 C-terminally. NOCIVA protein is the result of an alternative in-frame splicing of an intronic region with the preceding CIP2A protein coding sequence. NOCIVA protein thus comprises of 558 amino acids and contains a unique, immunogenic C-terminal 13 aa peptide tail that does not present sequence homology to any sequence in the human proteome. It is plausible that NOCIVA and CIP2A proteins have different cellular functions as CIP2A is known to mainly reside in the cytoplasm (Junttila et al. 2007), whereas NOCIVA was found to predominantly reside in the nucleus. However, NOCIVA maintained the ability to

dimerize and to bind to B56α, which would indicate that it functions similar to CIP2A as a PP2A inhibitor protein. Frustratingly, during this project I did not succeed in developing siRNA or CRISPR/Cas9 tools that would exclusively suppress NOCIVA. On the other hand, two NOCIVA specific RQ-PCR assays and antibodies were developed during this work and these can be further utilized in prospective NOCIVA studies. However, it is clear, that in the future appropriate functional models are required in order to unravel the precise role of NOCIVA mRNA and protein in the cells.

Relatively higher expression of *NOCIVA* than *CIP2A* was detected in AML and CML patient samples in our study. Interestingly, AML and CML are actually rather special in regard to *CIP2A* expression as they are the only cancer types where *CIP2A* has been reported to be underexpressed when compared to normal tissue (Lucas et al. 2011; Barragán et al. 2015). Exonic splicing silencer sequences (ESS), binding sites for hnRNPs and multiple splice factors were identified at the *NOCIVA* junction site by *in silico* predictions. ESSs are known to be linked to inhibition of pre-mRNA strands and promotion of exon skipping. On the other hand, a recent study reported expression changes of 13 hnRNPs affecting mRNA processing in AML (Alanazi et al. 2020) and out of these, hnRNP A1, A2B1, C were predicted to bind to the *NOCIVA* junction site. Moreover, the expression of hnRNP K (Gallardo et al. 2015), SRSF3 (Liu et al. 2012) and YB-1 (Liu et al. 2018) have been reported to be altered in AML, but also to take part in leukemia progression. All of these proteins were also predicted to bind to the *NOCIVA* junction site. Lastly, SRSF3 (Xiao et al. 2016) and YB-1 (Wei et al. 2012) have additionally been demonstrated to especially induce exon-inclusion during AS, the mechanism by which *NOCIVA* splicing occurs. Thus, in the future a detailed analysis of the role of these splicing modulators in the AS of *CIP2A* to *NOCIVA* is needed for a better understanding of the regulation of NOCIVA in myeloid cancers.

High *NOCIVA* expression at diagnosis associated with inferior outcome of both AML and CML patients in our study. In AML, *NOCIVA* expression was independent of the ELN-2010 genetic risk categories, indicating that the assessment of *NOCIVA* expression at diagnosis could provide clinically relevant additional value for pretreatment prediction of patient outcomes. In CML, high *NOCIVA* expression associated with inferior EFS and shorter FFP as well as with lower rates of CMR exclusively in imatinib treated patients. Thus, our results propose that first-line 2G TKI treatment might be needed to overcome the adverse effects caused by the high *NOCIVA* expression. Assessment of *NOCIVA* expression at CP CML diagnosis might thus help in treatment decisions between imatinib and 2G TKI as the first-line therapy. Together with the recent data that CIP2A protein expression, but not mRNA expression, predicts for resistance to both imatinib and 2G TKIs (Lucas et al. 2015), our data further highlight differential roles of *NOCIVA* and *CIP2A* in promoting

CML cell tolerance to TK inhibition (III, Figure 7). On the other hand, high CIP2A protein levels in CML indicate that sufficient amount of full-length *CIP2A* mRNA is transcribed to allow protein stabilization. We thus postulate that the therapy resistant CML cells are likely to harbor co-expression of both CIP2A and NOCIVA proteins to ensure maximal TKI resistance.

mRNA expression-based applications have recently entered clinical CML diagnostics. CML disease activity after TKI therapy can be nowadays followed up with an internationally standardized method that analyses the ratio of the expression levels of *BCR-ABL1* and *ABL1* transcripts (Radich et al. 2018). Recently, the *BCR-ABL1/ABL1* ratios on the international scale were also reported to predict the success of TFR attempt for patients with *de novo* CP CML (Nicolini et al. 2019). At the moment there are two technical platforms that allow standardized clinical assessment of mRNA expression levels for CML samples, the droplet digital PCR (ddPCR) (Nicolini et al. 2019) and the Cepheid GeneXpert qPCR cartridge system (Gerrard et al. 2016). However, currently there are no patient stratification markers guiding clinicians in the selection of first-line TKI treatment for optimal outcome in CML patient (Branford et al. 2019). Based on our nearly identical results from two independent clinical CML material, determination of *NOCIVA* mRNA levels from patients with *de novo* CP CML by ddPCR assay or Cepheid qPCR cartridge system, could provide significant support for clinicians in recognizing patients in need of frontline 2G TKI therapy. 2G TKI first-line treatment could further gain deeper molecular responses in these patients and subsequently enable more patients to attempt TKI discontinuation. Also, in AML, high *NOCIVA* expression at diagnosis could indicate a need for frontline therapy intensification. Thus, further validation of *NOCIVA* as a potent predictive biomarker in CML and AML is warranted.

7 Conclusions

The objective of this thesis was to identify novel cancer biomarkers among the PP2A inhibitor proteins in AML, CML and HNSCC. We made significant discoveries in our understanding of the role of ARPP19 and CIP2A as clinically useful biomarkers in these cancers. However, it has to be acknowledged that as most of the data presented are correlations with response, an important future perspective would be to determine the mechanisms that explain the involvement of CIP2A, ARPP19 and NOCIVA in the clinical events that they associate with. On the other hand, discovery of NOCIVA opens up a new horizon in the field of CIP2A isoforms and thus, further studies are warranted to explore the role of NOCIVA in both healthy tissues and cancer.

Based on the results of this thesis, the following key conclusions can be made:

1) CIP2A is an OCT4 target gene involved in HNSCC oncogenicity and radiation resistance. OCT4 positivity is linked to increased stemness of HNSCC tumors whereas CIP2A confers poor HNSCC patient survival. Future identification of OCT4 driven CIP2A target mechanisms may help in further understanding of the radioresistance in HNSCC as well as in development of novel radiosensitation therapies.

2) ARPP19 is a novel oncogene in AML that promotes CIP2A and MYC expression. *ARPP19* mRNA expression associates with patient relapse and disease activity in AML and could thus serve as a novel prognostic biomarker to detect an emerging relapse. In the future, ARPP19 mRNA levels could constitute a patient stratification strategy that guide patients with low *ARPP19* to standard chemotherapy, whereas patients with high *ARPP19* should be treated more intensively.

3) NOCIVA is a novel CIP2A splice variant that translates to a unique human protein. *NOCIVA* mRNA is a novel potential prognostic and predictive biomarker in AML and CML. Assessment of *NOCIVA* mRNA levels from patients at CP CML diagnosis, could provide significant support for clinicians in recognizing patients in need of frontline 2G TKI therapy.

Acknowledgements

This thesis work was carried out at the Faculty of Medicine, Department of Pathology, University of Turku and Turku Bioscience. I wish to acknowledge the director of Turku Bioscience, Professor John Eriksson, as well as the former director, Professor Riitta Lahesmaa, for providing an inspiring scientific atmosphere with top-notch research facilities. I warmly want to thank the technical staff and secretaries at the Turku Bioscience for making the Centre a great place to work at.

I want to express my immense gratitude to my supervisor, Professor Jukka Westermarck. Thank you for having me as a PhD student in your laboratory. It has truly been a privilege to be part of this amazing group of people and I have enjoyed working under your supervision throughout these years. I want to thank you for your warm guidance, support and encouragement. Especially I want to thank you for having faith in my projects even when my own faith wavered. Your enthusiasm and passion for science is a true inspiration.

I want to warmly acknowledge the members of my thesis committee, Assistant professor Noora Kotaja and Adjunct professor Veli Kairisto, for your interest to my research and for valuable advice and help during my doctoral training.

I would like to express my sincere gratitude to Turku Doctoral Programme of Molecular Medicine (TuDMM), for providing essential financial and educational support during my doctoral training. Especially I want to thank the director of TuDMM, Professor Kati Elima, and the coordinator Eeva Valve and former doctoral programme secretary Nina Widberg for their valuable work at running the doctoral programme. I would like to thank TuDMM for various enjoyable social events, but also fellow PhD students in TuDMM for an important source of peer support over the years. Additionally, I would like to thank chief academic officer of Medical Faculty, Outi Irjala, for the help with bureaucracy related to doctoral studies and the thesis book submission.

I am extremely grateful for the financial support that this work has received from the Turku Doctoral Programme of Molecular Medicine, Turku University Foundation, Cancer Foundation of Southwestern Finland, The Cultural Foundation of Southwestern Finland, Ida Montin Foundation, The Paulo Foundation, Cancer Foundation Finland, Emil Aaltonen Foundation, K Albin Johansson Foundation, The

Päivikki and Sakari Sohlberg Foundation, Cancer Foundation Väre, The Jenny and Antti Wihuri Foundation, The Sigrid Juselius Foundation, Turku University Hospital ERVA, The Finnish Concordia Fund and Business Finland TUTL.

Professor María D. Odero and Caroline Heckman are warmly acknowledged for critical evaluation of my thesis manuscript, and for giving insightful and constructive feedback which improved the quality of this thesis. I want to thank Professor Kimmo Porkka for accepting the invitation to be my opponent and at this date I hope we will have a lively scientific discussion.

I am truly thankful for all the co-authors and collaborators that I have had the opportunity work with during my doctoral training. Without your contribution, this work could not have been carried out. My sincere thanks go to Eliisa Löyttyniemi, Sami Ventelä, Urpu Salmenniemi, Karolina Pavic, Veli Kairisto, Maija Itälä-Remes, Taru Varila, Otto Kauko, Claire M. Lucas, Srikar Nagelli, Tea Ammunét, Veli-Matti Kähäri, Richard E Clark, Laura L. Elo, Venkata K. Bachanaboyina, Leni Kauko, Juho-Antti Mäkelä, Jarmo Kulmala, Leena Strauss, Olli Cárpen, Jorma Toppari, Reidar Grénman, Matti Varis, Piia von Lode and Marjut Borgenström.

It has been such a joy working in shared laboratory and office space with Westermarck, Jaakkola and Ivaska groups during these years. I am truly thankful for all the people I have shared the lab bench, the western blot table, the cell culture, the dark room and the inspiration over the years. Thank you all for the scientific and non-scientific conversations we have had. You all have made the roller-coaster of PhD smoother. I would like to warmly thank all the former and present members of Westermarck lab. Thank you Anni, Juha, Cristian, Inga, Minna, Tuuli, Karolina, Oxana, Xi, Joni, Umar, Srikar, Mukund, Anna L., Anna A., Tiina, Aman, Otto, Taru, Taina.

My deepest gratitude goes to people from Westermarck group who have guided me during my PhD studies. Thank you Anni for being a super important “big sister of science” in the beginning of my PhD journey. Thank you for guiding me into JW lab practice and for always having time for my silly questions. Otto, thank you for sharing your PP2A knowledge and for being a mentor in the beginning of my doctoral journey. You truly taught me a great deal about scientific thinking and also how important clear lab notes are. Thank you, Anna L., for sharing your good laboratory practice know-how and for showing me how to get things done efficiently and professionally. Juha, thank you for being you and thank you for teaching me the basics, and the not so basics, about buffers, proteins and Mordor language. Thank you, Aman, for your everlasting wisdom and positivity. It was such a joy working with you and the quality of your work is something to look up to. Xi, thank you for sitting next to me in the lab and in the office, and for joining me for the last phase of PhD journey. We have shared countless morning coffees, scientific and non-

scientific discussions over the past years. It has been such a joy working with you and to get to know you as a person.

My special thanks go to people from Westermarck group who have enabled and helped me during my doctoral training. Thank you Tiina, for making JW lab such a nice place to work at and for helping out with all the university and MELA bureaucracy. Taru, I want to especially thank you for being such a hardworking and excellent in all the experimental work that I have given you. You have truly been my right and left hand in the lab during the last couple of years. Without your contribution during the last couple of years, I would not have been able to get everything done in time. I also want to give an enormous thanks to Taina for your excellent technical assistance and for keeping the whole JW lab running. You are the core of Westermarck lab and the everyday work that you do is indispensable.

I am lucky to have many good friends in my life, and I want to warmly thank all of you for contributing immensely to the quality of life during my PhD journey. I would like to jointly thank "BioCityn misut" Inga, Petra, Heidi, Anni and Minna for sharing so much more than just work, for sharing countless fun moments and trips but also turning points in our lives over the years. I also want to acknowledge "TYKin tytöt" Pauliina, Anniina, Maija, Johanna and Soila, for following and supporting my path for the past 20 years already. It's such a privilege to have all of you in my life and I look forward for the next 20 years to come. Additionally, I want to thank my good friends Reetta and Johanna L. for always being there for me in the lab but especially outside the lab.

I want to deeply thank my family for always believing in me and for showing interest in my thesis studies. I want to thank my parents Liisa and Gerhardt for the unconditional love that continues to encourage me in everything I do. I want to thank my parents, my brother Mika and Antti's mom Päivikki, for all the support and love you have shown throughout these years.

Lastly, I want to thank my husband Antti for the past 15 years and for the countless years to come. You are the love of my life. Thank you for always supporting, loving and encouraging me in everything I do. I would not be here today if it wasn't for your support. Very lastly, I want to thank our sons, Kasperi and Rasmus, for showing me the essence of life and for teaching me more about life than I thought was possible. I love you both to the moon and back.

9.4.2021
Eleonora Mäkelä

References

Abelson, S., G. Collord, S. W. K. Ng, O. Weissbrod, N. Mendelson Cohen, E. Niemeyer, N. Barda, P. C. Zuzarte, L. Heisler, Y. Sundaravadanam, R. Luben, S. Hayat, T. T. Wang, Z. Zhao, I. Cirlan, T. J. Pugh, D. Soave, K. Ng, C. Latimer, C. Hardy, K. Raine, D. Jones, D. Hoult, A. Britten, J. D. McPherson, M. Johansson, F. Mbabaali, J. Eagles, J. K. Miller, D. Pasternack, L. Timms, P. Krzyzanowski, P. Awadalla, R. Costa, E. Segal, S. V. Bratman, P. Beer, S. Behjati, I. Martincorena, J. C. Y. Wang, K. M. Bowles, J. R. Quirós, A. Karakatsani, C. La Vecchia, A. Trichopoulou, E. Salamanca-Fernández, J. M. Huerta, A. Barricarte, R. C. Travis, R. Tumino, G. Masala, H. Boeing, S. Panico, R. Kaaks, A. Krämer, S. Sieri, E. Riboli, P. Vineis, M. Foll, J. McKay, S. Polidoro, N. Sala, K. T. Khaw, R. Vermeulen, P. J. Campbell, E. Papaemmanuil, M. D. Minden, A. Tanay, R. D. Balicer, N. J. Wareham, M. Gerstung, J. E. Dick, P. Brennan, G. S. Vassiliou, and L. I. Shlush. 2018. "Prediction of acute myeloid leukaemia risk in healthy individuals." *Nature* 559 (7714):400–404. doi: 10.1038/s41586-018-0317-6.

Abraham, S. A., L. E. Hopcroft, E. Carrick, M. E. Drotar, K. Dunn, A. J. Williamson, K. Korfi, P. Baquero, L. E. Park, M. T. Scott, F. Pellicano, A. Pierce, M. Copland, C. Nourse, S. M. Grimmond, D. Vetrie, A. D. Whetton, and T. L. Holyoake. 2016. "Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells." *Nature* 534 (7607):341–6. doi: 10.1038/nature18288.

Adamia, S., B. Haibe-Kains, P. M. Pilarski, M. Bar-Natan, S. Pevzner, H. Avet-Loiseau, L. Lode, S. Verselis, E. A. Fox, J. Burke, I. Galinsky, I. Dagogo-Jack, M. Wadleigh, D. P. Steensma, G. Motyckova, D. J. Deangelo, J. Quackenbush, R. Stone, and J. D. Griffin. 2014. "A genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets." *Clin Cancer Res* 20 (5):1135–45. doi: 10.1158/1078-0432.CCR-13-0956.

Agarwal, A., R. J. MacKenzie, R. Pippa, C. A. Eide, J. Oddo, J. W. Tyner, R. Sears, M. P. Vitek, M. D. Odero, D. J. Christensen, and B. J. Druker. 2014. "Antagonism of SET using OP449 enhances the efficacy of tyrosine kinase inhibitors and overcomes drug resistance in myeloid leukemia." *Clin Cancer Res* 20 (8):2092–103. doi: 10.1158/1078-0432.CCR-13-2575.

Agarwal, S., A. Kowalski, M. Schiffer, J. Zhao, J. P. Bewersdorf, and A. M. Zeidan. 2021. "Venetoclax for the treatment of elderly or chemotherapy-ineligible patients with acute myeloid leukemia: a step in the right direction or a game changer?" *Expert Rev Hematol*. doi: 10.1080/17474086.2021.1876559.

Al-Hajj, M., M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, and M. F. Clarke. 2003. "Prospective identification of tumorigenic breast cancer cells." *Proc Natl Acad Sci U S A* 100 (7):3983–8. doi: 10.1073/pnas.0530291100.

Al-Murrani, S. W., J. R. Woodgett, and Z. Damuni. 1999. "Expression of I2PP2A, an inhibitor of protein phosphatase 2A, induces c-Jun and AP-1 activity." *Biochem J* 341 (Pt 2):293–8.

Alanazi, B., C. R. Munje, N. Rastogi, A. J. K. Williamson, S. Taylor, P. S. Hole, M. Hodges, M. Doyle, S. Baker, A. F. Gilkes, S. Knapper, A. Pierce, A. D. Whetton, R. L. Darley, and A. Tonks. 2020.

"Integrated nuclear proteomics and transcriptomics identifies S100A4 as a therapeutic target in acute myeloid leukemia." *Leukemia* 34 (2):427–440. doi: 10.1038/s41375-019-0596-4.

Albers, A. E., C. Chen, B. Köberle, X. Qian, J. P. Klussmann, B. Wollenberg, and A. M. Kaufmann. 2012. "Stem cells in squamous head and neck cancer." *Crit Rev Oncol Hematol* 81 (3):224–40. doi: 10.1016/j.critrevonc.2011.03.004.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25 (17):3389–402.

Amin, M. B., F. L. Greene, S. B. Edge, C. C. Compton, J. E. Gershenwald, R. K. Brookland, L. Meyer, D. M. Gress, D. R. Byrd, and D. P. Winchester. 2017. "The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging." *CA Cancer J Clin* 67 (2):93–99. doi: 10.3322/caac.21388.

Andrade, E. C., V. Musante, A. Horiuchi, H. Matsuzaki, A. H. Brody, T. Wu, P. Greengard, J. R. Taylor, and A. C. Nairn. 2017. "ARPP-16 Is a Striatal-Enriched Inhibitor of Protein Phosphatase 2A Regulated by Microtubule-Associated Serine/Threonine Kinase 3 (Mast 3 Kinase)." *J Neurosci* 37 (10):2709–2722. doi: 10.1523/JNEUROSCI.4559-15.2017.

Andrews, C. N., and J. Lipton. 2019. "The true value of second-generation TKIs as first-line therapy in chronic myeloid leukaemia." *Lancet Haematol* 6 (8):e385–e386. doi: 10.1016/S2352-3026(19)30112-7.

Arber, D. A., A. Orazi, R. Hasserjian, J. Thiele, M. J. Borowitz, M. M. Le Beau, C. D. Bloomfield, M. Cazzola, and J. W. Vardiman. 2016. "The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia." *Blood* 127 (20):2391–405. doi: 10.1182/blood-2016-03-643544.

Arriazu, E., R. Pippa, and M. D. Odero. 2016. "Protein Phosphatase 2A as a Therapeutic Target in Acute Myeloid Leukemia." *Front Oncol* 6:78. doi: 10.3389/fonc.2016.00078.

Ausoni, S., P. Boscolo-Rizzo, B. Singh, M. C. Da Mosto, G. Spinato, G. Tirelli, R. Spinato, and G. Azzarello. 2016. "Targeting cellular and molecular drivers of head and neck squamous cell carcinoma: current options and emerging perspectives." *Cancer Metastasis Rev* 35 (3):413–26. doi: 10.1007/s10555-016-9625-1.

Banerjee, R., N. Russo, M. Liu, V. Basrur, E. Bellile, N. Palanisamy, C. S. Scanlon, E. van Tubergen, R. C. Inglehart, T. Metwally, R. S. Mani, A. Yocum, M. K. Nyati, R. M. Castilho, S. Varambally, A. M. Chinnaiyan, and N. J. D'Silva. 2014. "TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer." *Nat Commun* 5:4527. doi: 10.1038/ncomms5527.

Barragán, E., M. C. Chillón, R. Castelló-Cros, N. Marcotegui, M. I. Prieto, M. Hoyos, R. Pippa, M. Llop, A. Etxabe, J. Cervera, G. Rodríguez, I. Buño, J. Rifón, J. Sierra, M. González, M. J. Calasanz, M. A. Sanz, and M. D. Odero. 2015. "CIP2A high expression is a poor prognostic factor in normal karyotype acute myeloid leukemia." *Haematologica* 100 (5):e183–5. doi: 10.3324/haematol.2014.118117.

Batlle, E., and H. Clevers. 2017. "Cancer stem cells revisited." *Nat Med* 23 (10):1124–1134. doi: 10.1038/nm.4409.

Bialojan, C., and A. Takai. 1988. "Inhibitory effect of a marine-sponge toxin, okadaic acid, on protein phosphatases. Specificity and kinetics." *Biochem J* 256 (1):283–90. doi: 10.1042/bj2560283.

Birkman, E. M., A. Elzagheid, T. Jokilehto, T. Avoranta, E. Korkeila, J. Kulmala, K. Syrjänen, J. Westermarck, and J. Sundström. 2018. "Protein phosphatase 2A (PP2A) inhibitor CIP2A indicates resistance to radiotherapy in rectal cancer." *Cancer Med* 7 (3):698–706. doi: 10.1002/cam4.1361.

Bonnet, D., and J. E. Dick. 1997. "Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell." *Nat Med* 3 (7):730–7. doi: 10.1038/nm0797-730.

Branford, S., D. D. H. Kim, J. F. Apperley, C. A. Eide, S. Mustjoki, S. T. Ong, G. Nteliopoulos, T. Ernst, C. Chuah, C. Gambacorti-Passerini, M. J. Mauro, B. J. Druker, D. W. Kim, F. X. Mahon, J. Cortes, J. P. Radich, A. Hochhaus, T. P. Hughes, and International CML Foundation Genomics Alliance. 2019. "Laying the foundation for genomically-based risk assessment in chronic myeloid leukemia." *Leukemia* 33 (8):1835–1850. doi: 10.1038/s41375-019-0512-y.

Branford, S., P. Wang, D. T. Yeung, D. Thomson, A. Purins, C. Wadham, N. H. Shahrin, J. E. Marum, N. Nataren, W. T. Parker, J. Geoghegan, J. Feng, N. Shanmuganathan, M. C. Mueller, C. Dietz, D. Stangl, Z. Donaldson, H. Altamura, J. Georgievski, J. Braley, A. Brown, C. Hahn, I. Walker, S. H. Kim, S. Y. Choi, S. H. Park, D. W. Kim, D. L. White, A. S. M. Yong, D. M. Ross, H. S. Scott, A. W. Schreiber, and T. P. Hughes. 2018. "Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease." *Blood* 132 (9):948–961. doi: 10.1182/blood-2018-02-832253.

Braun, T. P., C. A. Eide, and B. J. Druker. 2020. "Response and Resistance to BCR-ABL1-Targeted Therapies." *Cancer Cell* 37 (4):530–542. doi: 10.1016/j.ccell.2020.03.006.

Bray, F., J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. 2018. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA Cancer J Clin* 68 (6):394–424. doi: 10.3322/caac.21492.

Bruchova-Votavova, H., D. Yoon, and J. T. Prchal. 2010. "miR-451 enhances erythroid differentiation in K562 cells." *Leuk Lymphoma* 51 (4):686–93. doi: 10.3109/10428191003629362.

Bryja, V., S. Bonilla, and E. Arenas. 2006. "Derivation of mouse embryonic stem cells." *Nat Protoc* 1 (4):2082–7. doi: 10.1038/nprot.2006.355.

Böckelman, C., J. Hagström, L. K. Mäkinen, H. Keski-Säntti, V. Häyry, J. Lundin, T. Atula, A. Ristimäki, and C. Haglund. 2011. "High CIP2A immunoreactivity is an independent prognostic indicator in early-stage tongue cancer." *Br J Cancer* 104 (12):1890–5. doi: 10.1038/bjc.2011.167.

Böckelman, C., H. Lassus, A. Hemmes, A. Leminen, J. Westermarck, C. Haglund, R. Bützow, and A. Ristimäki. 2011. "Prognostic role of CIP2A expression in serous ovarian cancer." *Br J Cancer* 105 (7):989–95. doi: 10.1038/bjc.2011.346.

Cabrera, M. C., R. E. Hollingsworth, and E. M. Hurt. 2015. "Cancer stem cell plasticity and tumor hierarchy." *World J Stem Cells* 7 (1):27–36. doi: 10.4252/wjsc.v7.i1.27.

Califf, R. M. 2018. "Biomarker definitions and their applications." *Exp Biol Med (Maywood)* 243 (3):213–221. doi: 10.1177/1535370217750088.

Cerami, E., J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, and N. Schultz. 2012. "The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data." *Cancer Discov* 2 (5):401–4. doi: 10.1158/2159-8290.CD-12-0095.

Chang, L. S., S. Pan, M. M. Pater, and G. Di Mayorca. 1985. "Differential requirement for SV40 early genes in immortalization and transformation of primary rat and human embryonic cells." *Virology* 146 (2):246–61. doi: 10.1016/0042-6822(85)90008-x.

Chen, J., B. L. Martin, and D. L. Brautigan. 1992. "Regulation of protein serine-threonine phosphatase type-2A by tyrosine phosphorylation." *Science* 257 (5074):1261–4. doi: 10.1126/science.1325671.

Chen, K. F., C. Y. Liu, Y. C. Lin, H. C. Yu, T. H. Liu, D. R. Hou, P. J. Chen, and A. L. Cheng. 2010. "CIP2A mediates effects of bortezomib on phospho-Akt and apoptosis in hepatocellular carcinoma cells." *Oncogene* 29 (47):6257–66. doi: 10.1038/onc.2010.357.

Chen, M. J., J. E. Dixon, and G. Manning. 2017. "Genomics and evolution of protein phosphatases." *Sci Signal* 10 (474). doi: 10.1126/scisignal.aag1796.

Chen, Y. C., H. S. Hsu, Y. W. Chen, T. H. Tsai, C. K. How, C. Y. Wang, S. C. Hung, Y. L. Chang, M. L. Tsai, Y. Y. Lee, H. H. Ku, and S. H. Chiou. 2008. "Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells." *PLoS One* 3 (7):e2637. doi: 10.1371/journal.pone.0002637.

Chen, Y. L., M. H. Kuo, P. Y. Lin, W. L. Chuang, C. C. Hsu, P. Y. Chu, C. H. Lee, T. H. Huang, Y. W. Leu, and S. H. Hsiao. 2013. "ENSA expression correlates with attenuated tumor propagation in liver cancer." *Biochem Biophys Res Commun* 442 (1-2):56–61. doi: 10.1016/j.bbrc.2013.10.165.

Chiou, S. H., C. C. Yu, C. Y. Huang, S. C. Lin, C. J. Liu, T. H. Tsai, S. H. Chou, C. S. Chien, H. H. Ku, and J. F. Lo. 2008. "Positive correlations of Oct-4 and Nanog in oral cancer stem-like cells and high-grade oral squamous cell carcinoma." *Clin Cancer Res* 14 (13):4085–95. doi: 10.1158/1078-0432.CCR-07-4404.

Choi, Y. A., J. S. Park, M. Y. Park, K. S. Oh, M. S. Lee, J. S. Lim, K. I. Kim, K. Y. Kim, J. Kwon, D. Y. Yoon, E. Y. Moon, and Y. Yang. 2011. "Increase in CIP2A expression is associated with doxorubicin resistance." *FEBS Lett* 585 (5):755–60. doi: 10.1016/j.febslet.2011.01.018.

Chow, L. Q. M. 2020. "Head and Neck Cancer. Reply." *N Engl J Med* 382 (20):e57. doi: 10.1056/NEJMc2001370.

Christensen, D. J., Y. Chen, J. Oddo, K. M. Matta, J. Neil, E. D. Davis, A. D. Volkheimer, M. C. Lanasa, D. R. Friedman, B. K. Goodman, J. P. Gockerman, L. F. Diehl, C. M. de Castro, J. O. Moore, M. P. Vitek, and J. B. Weinberg. 2011. "SET oncoprotein overexpression in B-cell chronic lymphocytic leukemia and non-Hodgkin lymphoma: a predictor of aggressive disease and a new treatment target." *Blood* 118 (15):4150–8. doi: 10.1182/blood-2011-04-351072.

Clark, A. T. 2007. "The stem cell identity of testicular cancer." *Stem Cell Rev* 3 (1):49–59. doi: 10.1007/s12015-007-0002-x.

Clarkson, B., J. Fried, A. Strife, Y. Sakai, K. Ota, and T. Okita. 1970. "Studies of cellular proliferation in human leukemia. 3. Behavior of leukemic cells in three adults with acute leukemia given continuous infusions of 3H-thymidine for 8 or 10 days." *Cancer* 25 (6):1237–60. doi: 10.1002/1097-0142(197006)25:6<1237::aid-cncr2820250602>3.0.co;2-7.

Clarkson, B., T. Ohkita, K. Ota, and J. Fried. 1967. "Studies of cellular proliferation in human leukemia. I. Estimation of growth rates of leukemic and normal hematopoietic cells in two adults with acute leukemia given single injections of tritiated thymidine." *J Clin Invest* 46 (4):506–29. doi: 10.1172/JCI105553.

Climente-González, H., E. Porta-Pardo, A. Godzik, and E. Eyras. 2017. "The Functional Impact of Alternative Splicing in Cancer." *Cell Rep* 20 (9):2215–2226. doi: 10.1016/j.celrep.2017.08.012.

Corces, M. R., J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, R. Majeti, and H. Y. Chang. 2016. "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution." *Nat Genet* 48 (10):1193–203. doi: 10.1038/ng.3646.

Cosgun, K. N., S. Rahmig, N. Mende, S. Reinke, I. Hauber, C. Schäfer, A. Petzold, H. Weisbach, G. Heidkamp, A. Purbojo, R. Cesnjevar, A. Platz, M. Bornhäuser, M. Schmitz, D. Dudziak, J. Hauber, J. Kirberg, and C. Waskow. 2014. "Kit regulates HSC engraftment across the human-mouse species barrier." *Cell Stem Cell* 15 (2):227–38. doi: 10.1016/j.stem.2014.06.001.

Cristóbal, I., L. Garcia-Orti, C. Cirauqui, X. Cortes-Lavaud, M. A. García-Sánchez, M. J. Calasanz, and M. D. Odero. 2012. "Overexpression of SET is a recurrent event associated with poor outcome and contributes to protein phosphatase 2A inhibition in acute myeloid leukemia." *Haematologica* 97 (4):543–50. doi: 10.3324/haematol.2011.050542.

de Necochea-Campion, R., G. P. Shouse, Q. Zhou, S. Mirshahidi, and C. S. Chen. 2016. "Aberrant splicing and drug resistance in AML." *J Hematol Oncol* 9 (1):85. doi: 10.1186/s13045-016-0315-9.

Desai, P., N. Mencia-Trinchant, O. Savenkov, M. S. Simon, G. Cheang, S. Lee, M. Samuel, E. K. Ritchie, M. L. Guzman, K. V. Ballman, G. J. Roboz, and D. C. Hassane. 2018. "Somatic mutations precede acute myeloid leukemia years before diagnosis." *Nat Med* 24 (7):1015–1023. doi: 10.1038/s41591-018-0081-z.

Desmet, F. O., D. Hamroun, M. Lalande, G. Collod-Béroud, M. Claustres, and C. Béroud. 2009. "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." *Nucleic Acids Res* 37 (9):e67. doi: 10.1093/nar/gkp215.

Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson, and J. F. DiPersio. 2012. "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing." *Nature* 481 (7382):506–10. doi: 10.1038/nature10738.

Dirkse, A., A. Golebiewska, T. Buder, P. V. Nazarov, A. Muller, S. Poovathingal, N. H. C. Brons, S. Leite, N. Sauvageot, D. Sarkisjan, M. Seyfrid, S. Fritah, D. Stieber, A. Michelucci, F. Hertel, C. Herold-Mende, F. Azuaje, A. Skupin, R. Bjerkvig, A. Deutsch, A. Voss-Böhme, and S. P. Niclou. 2019. "Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment." *Nat Commun* 10 (1):1787. doi: 10.1038/s41467-019-09853-z.

Duan, G., X. Li, and M. Köhn. 2015. "The human DEPhOsphorylation database DEPOD: a 2015 update." *Nucleic Acids Res* 43 (Database issue):D531–5. doi: 10.1093/nar/gku1009.

Döhner, H., E. Estey, D. Grimwade, S. Amadori, F. R. Appelbaum, T. Büchner, H. Dombret, B. L. Ebert, P. Fenaux, R. A. Larson, R. L. Levine, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. Sanz, J. Sierra, M. S. Tallman, H. F. Tien, A. H. Wei, B. Löwenberg, and C. D. Bloomfield. 2017. "Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel." *Blood* 129 (4):424–447. doi: 10.1182/blood-2016-08-733196.

Döhner, H., E. H. Estey, S. Amadori, F. R. Appelbaum, T. Büchner, A. K. Burnett, H. Dombret, P. Fenaux, D. Grimwade, R. A. Larson, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. A. Sanz, J. Sierra, M. S. Tallman, B. Löwenberg, C. D. Bloomfield, and European LeukemiaNet. 2010. "Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet." *Blood* 115 (3):453–74. doi: 10.1182/blood-2009-07-235358.

Döhner, H., D. J. Weisdorf, and C. D. Bloomfield. 2015. "Acute Myeloid Leukemia." *N Engl J Med* 373 (12):1136–52. doi: 10.1056/NEJMra1406184.

Eichhorn, P. J., M. P. Creyghton, and R. Bernards. 2009. "Protein phosphatase 2A regulatory subunits and cancer." *Biochim Biophys Acta* 1795 (1):1–15. doi: 10.1016/j.bbcan.2008.05.005.

Enjoji, S., R. Yabe, S. Tsuji, K. Yoshimura, H. Kawasaki, M. Sakurai, Y. Sakai, H. Takenouchi, S. Yoshino, S. Hazama, H. Nagano, H. Oshima, M. Oshima, M. P. Vitek, T. Matsuura, Y. Hippo, T. Usui, T. Ohama, and K. Sato. 2018. "Stemness Is Enhanced in Gastric Cancer by a SET/PP2A/E2F1 Axis." *Mol Cancer Res* 16 (3):554–563. doi: 10.1158/1541-7786.MCR-17-0393.

Eppert, K., K. Takenaka, E. R. Lechman, L. Waldron, B. Nilsson, P. van Galen, K. H. Metzeler, A. Poeppl, V. Ling, J. Beyene, A. J. Canty, J. S. Danska, S. K. Bohlander, C. Buske, M. D. Minden,

T. R. Golub, I. Jurisica, B. L. Ebert, and J. E. Dick. 2011. "Stem cell gene expression programs influence clinical outcome in human leukemia." *Nat Med* 17 (9):1086–93. doi: 10.1038/nm.2415.

Faderl, S., M. Talpaz, Z. Estrov, S. O'Brien, R. Kurzrock, and H. M. Kantarjian. 1999. "The biology of chronic myeloid leukemia." *N Engl J Med* 341 (3):164–72. doi: 10.1056/NEJM199907153410306.

Fan, Z., M. Li, X. Chen, J. Wang, X. Liang, H. Wang, Z. Wang, B. Cheng, and J. Xia. 2017. "Prognostic Value of Cancer Stem Cell Markers in Head and Neck Squamous Cell Carcinoma: a Meta-analysis." *Sci Rep* 7:43008. doi: 10.1038/srep43008.

Farnebo, L., F. Jerhammar, R. Ceder, R. C. Grafström, L. Vainikka, L. Thunell, R. Grénman, A. C. Johansson, and K. Roberg. 2011. "Combining factors on protein and gene level to predict radioresponse in head and neck cancer cell lines." *J Oral Pathol Med* 40 (10):739–46. doi: 10.1111/j.1600-0714.2011.01036.x.

Ferlay, J., M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray. 2019. "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods." *Int J Cancer* 144 (8):1941–1953. doi: 10.1002/ijc.31937.

Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. Coebergh, H. Comber, D. Forman, and F. Bray. 2013. "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012." *Eur J Cancer* 49 (6):1374–403. doi: 10.1016/j.ejca.2012.12.027.

Fialkow, P. J. 1974. "The origin and development of human tumors studied with cell markers." *N Engl J Med* 291 (1):26–35. doi: 10.1056/NEJM197407042910109.

Fleuren, E. D., L. Zhang, J. Wu, and R. J. Daly. 2016. "The kinome 'at large' in cancer." *Nat Rev Cancer* 16 (2):83–98. doi: 10.1038/nrc.2015.18.

Fowle, H., Z. Zhao, and X. Graña. 2019. "PP2A holoenzymes, substrate specificity driving cellular functions and deregulation in cancer." *Adv Cancer Res* 144:55–93. doi: 10.1016/bs.acr.2019.03.009.

Furness, S., A. M. Glenny, H. V. Worthington, S. Pavitt, R. Oliver, J. E. Clarkson, M. Macluskey, K. K. Chan, and D. I. Conway. 2011. "Interventions for the treatment of oral cavity and oropharyngeal cancer: chemotherapy." *Cochrane Database Syst Rev* (4):CD006386. doi: 10.1002/14651858.CD006386.pub3.

Gallardo, M., H. J. Lee, X. Zhang, C. Bueso-Ramos, L. R. Pageon, M. McArthur, A. Multani, A. Nazha, T. Manshouri, J. Parker-Thornburg, I. Rapado, A. Quintas-Cardama, S. M. Kornblau, J. Martinez-Lopez, and S. M. Post. 2015. "hnRNP K Is a Haploinsufficient Tumor Suppressor that Regulates Proliferation and Differentiation Programs in Hematologic Malignancies." *Cancer Cell* 28 (4):486–499. doi: 10.1016/j.ccell.2015.09.001.

Gambacorti-Passerini, C., L. Antolini, F. X. Mahon, F. Guilhot, M. Deininger, C. Fava, A. Nagler, C. M. Della Casa, E. Morra, E. Abruzzese, A. D'Emilio, F. Stagno, P. le Coutre, R. Hurtado-Monroy, V. Santini, B. Martino, F. Pane, A. Piccin, P. Giraldo, S. Assouline, M. A. Durosinmi, O. Leeksma, E. M. Pogliani, M. Puttini, E. Jang, J. Reiffers, Piazza, M. G. Valsecchi, and D. W. Kim. 2011. "Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib." *J Natl Cancer Inst* 103 (7):553–61. doi: 10.1093/jnci/djr060.

Gao, X., C. Lu, C. Chen, K. Sun, Q. Liang, J. Shuai, X. Wang, and Y. Xu. 2020. "ARPP-19 Mediates Herceptin Resistance via Regulation of CD44 in Gastric Cancer." *Onco Targets Ther* 13:6629–6643. doi: 10.2147/OTT.S253841.

Gavosto, F., A. Pileri, V. Gabutti, and P. Masera. 1967. "Non-self-maintaining kinetics of proliferating blasts in human acute leukaemia." *Nature* 216 (5111):188–9. doi: 10.1038/216188a0.

Gentry, S. N., and T. L. Jackson. 2013. "A mathematical model of cancer stem cell driven tumor initiation: implications of niche size and loss of homeostatic regulatory mechanisms." *PLoS One* 8 (8):e71128. doi: 10.1371/journal.pone.0071128.

Gerrard, G., H. E. Foong, K. Mudge, M. Alikian, J. F. Apperley, and L. Foroni. 2016. "Cepheid xpert monitor platform for the confirmation of BCR-ABL1 IS conversion factors for the molecular monitoring of chronic myeloid leukaemia." *Leuk Res* 49:47–50. doi: 10.1016/j.leukres.2016.08.007.

Gerstung, M., E. Papaemmanuil, I. Martincorena, L. Bullinger, V. I. Gaidzik, P. Paschka, M. Heuser, F. Thol, N. Bolli, P. Ganly, A. Ganser, U. McDermott, K. Döhner, R. F. Schlenk, H. Döhner, and P. J. Campbell. 2017. "Precision oncology for acute myeloid leukemia using a knowledge bank approach." *Nat Genet* 49 (3):332–340. doi: 10.1038/ng.3756.

Gharbi-Ayachi, A., J. C. Labbé, A. Burgess, S. Vigneron, J. M. Strub, E. Brioudes, A. Van-Dorsselaer, A. Castro, and T. Lorca. 2010. "The substrate of Greatwall kinase, Arpp19, controls mitosis by inhibiting protein phosphatase 2A." *Science* 330 (6011):1673–7. doi: 10.1126/science.1197048.

Ghisolfi, L., A. C. Keates, X. Hu, D. K. Lee, and C. J. Li. 2012. "Ionizing radiation induces stemness in cancer cells." *PLoS One* 7 (8):e43628. doi: 10.1371/journal.pone.0043628.

Girault, J. A., A. Horiuchi, E. L. Gustafson, N. L. Rosen, and P. Greengard. 1990. "Differential expression of ARPP-16 and ARPP-19, two highly related cAMP-regulated phosphoproteins, one of which is specifically associated with dopamine-innervated brain regions." *J Neurosci* 10 (4):1124–33.

Goldman, Mary, Brian Craft, Angela N Brooks, Jingchun Zhu, and David Haussler. 2018. The UCSC Xena Platform for cancer genomics data visualization and interpretation. bioRxiv.

Gong, Y., W. Wu, X. Zou, F. Liu, T. Wei, and J. Zhu. 2018. "MiR-26a inhibits thyroid cancer cell proliferation by targeting ARPP19." *Am J Cancer Res* 8 (6):1030–1039.

Goossens, N., S. Nakagawa, X. Sun, and Y. Hoshida. 2015. "Cancer biomarker discovery and validation." *Transl Cancer Res* 4 (3):256–269. doi: 10.3978/j.issn.2218-676X.2015.06.04.

Grech, G., J. Pollacco, M. Portelli, K. Sacco, S. Baldacchino, J. Grixti, and C. Saliba. 2014. "Expression of different functional isoforms in haematopoiesis." *Int J Hematol* 99 (1):4–11. doi: 10.1007/s12185-013-1477-7.

Griffin, J. D., and B. Löwenberg. 1986. "Clonogenic cells in acute myeloblastic leukemia." *Blood* 68 (6):1185–95.

Grimwade, D., A. Ivey, and B. J. Huntly. 2016. "Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance." *Blood* 127 (1):29–41. doi: 10.1182/blood-2015-07-604496.

Guenebeaud, C., D. Goldschneider, M. Castets, C. Guix, G. Chazot, C. Delloye-Bourgeois, A. Eisenberg-Lerner, G. Shohat, M. Zhang, V. Laudet, A. Kimchi, A. Bernet, and P. Mehlen. 2010. "The dependence receptor UNC5H2/B triggers apoptosis via PP2A-mediated dephosphorylation of DAP kinase." *Mol Cell* 40 (6):863–76. doi: 10.1016/j.molcel.2010.11.021.

Gupta, P. B., I. Pastushenko, A. Skibinski, C. Blanpain, and C. Kuperwasser. 2019. "Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance." *Cell Stem Cell* 24 (1):65–78. doi: 10.1016/j.stem.2018.11.011.

Hached, K., P. Goguet, S. Charrasse, S. Vigneron, M. P. Sacristan, T. Lorca, and A. Castro. 2019. "ENSA and ARPP19 differentially control cell cycle progression and development." *J Cell Biol* 218 (2):541–558. doi: 10.1083/jcb.201708105.

Haesen, D., L. Abbasi Asbagh, R. Derua, A. Hubert, S. Schrauwen, Y. Hoorne, F. Amant, E. Waelkens, A. Sablina, and V. Janssens. 2016. "Recurrent PPP2R1A Mutations in Uterine Cancer Act through a Dominant-Negative Mechanism to Promote Malignant Cell Growth." *Cancer Res* 76 (19):5719–5731. doi: 10.1158/0008-5472.CAN-15-3342.

Haesen, D., W. Sents, K. Lemaire, Y. Hoorne, and V. Janssens. 2014. "The Basic Biology of PP2A in Hematologic Cells and Malignancies." *Front Oncol* 4:347. doi: 10.3389/fonc.2014.00347.

Hahn, W. C., C. M. Counter, A. S. Lundberg, R. L. Beijersbergen, M. W. Brooks, and R. A. Weinberg. 1999. "Creation of human tumour cells with defined genetic elements." *Nature* 400 (6743):464–8. doi: 10.1038/22780.

Hahn, W. C., S. K. Dessain, M. W. Brooks, J. E. King, B. Elenbaas, D. M. Sabatini, J. A. DeCaprio, and R. A. Weinberg. 2002. "Enumeration of the simian virus 40 early region elements necessary for human cell transformation." *Mol Cell Biol* 22 (7):2111–23.

Hahn, W. C., and R. A. Weinberg. 2002. "Rules for making human tumor cells." *N Engl J Med* 347 (20):1593–603. doi: 10.1056/NEJMra021902.

Han, J., T. Fujisawa, S. R. Husain, and R. K. Puri. 2014. "Identification and characterization of cancer stem cells in human head and neck squamous cell carcinoma." *BMC Cancer* 14:173. doi: 10.1186/1471-2407-14-173.

Hanahan, D., and R. A. Weinberg. 2000. "The hallmarks of cancer." *Cell* 100 (1):57-70.

Hanahan, D., and R. A. Weinberg. 2011. "Hallmarks of cancer: the next generation." *Cell* 144 (5):646–74. doi: 10.1016/j.cell.2011.02.013.

Harari, P. M., D. L. Wheeler, and J. R. Grandis. 2009. "Molecular target approaches in head and neck cancer: epidermal growth factor receptor and beyond." *Semin Radiat Oncol* 19 (1):63–8. doi: 10.1016/j.semradonc.2008.09.009.

Herold, T., M. Rothenberg-Thurley, V. V. Grunwald, H. Janke, D. Goerlich, M. C. Sauerland, N. P. Konstandin, A. Dufour, S. Schneider, M. Neusser, B. Ksienzyk, P. A. Greif, M. Subklewe, A. Faldum, S. K. Bohlander, J. Braess, B. Wörmann, U. Krug, W. E. Berdel, W. Hiddemann, K. Spiekermann, and K. H. Metzeler. 2020. "Validation and refinement of the revised 2017 European LeukemiaNet genetic risk stratification of acute myeloid leukemia." *Leukemia*. doi: 10.1038/s41375-020-0806-0.

Hochhaus, A., M. Baccarani, R. T. Silver, C. Schiffer, J. F. Apperley, F. Cervantes, R. E. Clark, J. E. Cortes, M. W. Deininger, F. Guilhot, H. Hjorth-Hansen, T. P. Hughes, J. J. W.M Janssen, H. M. Kantarjian, D. W. Kim, R. A. Larson, J. H. Lipton, F. X. Mahon, J. Mayer, F. Nicolini, D. Niederwieser, F. Pane, J. P. Radich, D. Rea, J. Richter, G. Rosti, P. Rousselot, G. Saglio, S. Saußele, S. Soverini, J. L. Steegmann, A. Turkina, A. Zaritskey, and R. Hehlmann. 2020. "European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia." *Leukemia* 34 (4):966–984. doi: 10.1038/s41375-020-0776-2.

Hochhaus, A., M. Breccia, G. Saglio, V. García-Gutiérrez, D. Réa, J. Janssen, and J. Apperley. 2020. "Expert opinion-management of chronic myeloid leukemia after resistance to second-generation tyrosine kinase inhibitors." *Leukemia* 34 (6):1495–1502. doi: 10.1038/s41375-020-0842-9.

Hochhaus, A., R. A. Larson, F. Guilhot, J. P. Radich, S. Branford, T. P. Hughes, M. Baccarani, M. W. Deininger, F. Cervantes, S. Fujihara, C. E. Ortmann, H. D. Menssen, H. Kantarjian, S. G. O'Brien, B. J. Druker, and IRIS Investigators. 2017. "Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia." *N Engl J Med* 376 (10):917–927. doi: 10.1056/NEJMoa1609324.

Holyoake, T., X. Jiang, C. Eaves, and A. Eaves. 1999. "Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia." *Blood* 94 (6):2056–64.

Hourigan, C. S., R. P. Gale, N. J. Gormley, G. J. Ossenkoppele, and R. B. Walter. 2017. "Measurable residual disease testing in acute myeloid leukaemia." *Leukemia* 31 (7):1482–1490. doi: 10.1038/leu.2017.113.

Hsieh, J. C., H. M. Wang, M. H. Wu, K. P. Chang, P. H. Chang, C. T. Liao, and C. T. Liau. 2019. "Review of emerging biomarkers in head and neck squamous cell carcinoma in the era of immunotherapy and targeted therapy." *Head Neck* 41 Suppl 1:19–45. doi: 10.1002/hed.25932.

Hunter, T. 1995. "Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling." *Cell* 80 (2):225–36. doi: 10.1016/0092-8674(95)90405-0.

Hutchinson, M. N. D., M. Mierzwa, and N. J. D'Silva. 2020. "Radiation resistance in head and neck squamous cell carcinoma: dire need for an appropriate sensitizer." *Oncogene* 39 (18):3638–3649. doi: 10.1038/s41388-020-1250-3.

Intlekofer, A. M., and L. W. S. Finley. 2019. "Metabolic signatures of cancer cells and stem cells." *Nat Metab* 1 (2):177–188. doi: 10.1038/s42255-019-0032-0.

Jan, M., T. M. Snyder, M. R. Corces-Zimmerman, P. Vyas, I. L. Weissman, S. R. Quake, and R. Majeti. 2012. "Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia." *Sci Transl Med* 4 (149):149ra118. doi: 10.1126/scitranslmed.3004315.

Janssens, V., and J. Goris. 2001. "Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling." *Biochem J* 353 (Pt 3):417–39.

Jiang, T., B. Zhao, X. Li, and J. Wan. 2016. "ARPP-19 promotes proliferation and metastasis of human glioma." *Neuroreport* 27 (13):960–6. doi: 10.1097/WNR.0000000000000638.

Jung, N., B. Dai, A. J. Gentles, R. Majeti, and A. P. Feinberg. 2015. "An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis." *Nat Commun* 6:8489. doi: 10.1038/ncomms9489.

Junttila, M. R., P. Puustinen, M. Niemelä, R. Ahola, H. Arnold, T. Böttzauw, R. Ala-aho, C. Nielsen, J. Ivaska, Y. Taya, S. L. Lu, S. Lin, E. K. Chan, X. J. Wang, R. Grènman, J. Kast, T. Kallunki, R. Sears, V. M. Kähäri, and J. Westermarck. 2007. "CIP2A inhibits PP2A in human malignancies." *Cell* 130 (1):51–62. doi: 10.1016/j.cell.2007.04.044.

Kahles, A., K. V. Lehmann, N. C. Toussaint, M. Hüser, S. G. Stark, T. Sachsenberg, O. Stegle, O. Kohlbacher, C. Sander, G. Rätsch, and Cancer Genome Atlas Research Network. 2018. "Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients." *Cancer Cell* 34 (2):211–224.e6. doi: 10.1016/j.ccell.2018.07.001.

Kalla, C., M. O. Scheuermann, I. Kube, M. Schlotter, D. Mertens, H. Döhner, S. Stilgenbauer, and P. Lichter. 2007. "Analysis of 11q22-q23 deletion target genes in B-cell chronic lymphocytic leukaemia: evidence for a pathogenic role of NPAT, CUL5, and PPP2R1B." *Eur J Cancer* 43 (8):1328–35. doi: 10.1016/j.ejca.2007.02.005.

Kanatsu-Shinohara, M., K. Inoue, J. Lee, M. Yoshimoto, N. Ogonuki, H. Miki, S. Baba, T. Kato, Y. Kazuki, S. Toyokuni, M. Toyoshima, O. Niwa, M. Oshimura, T. Heike, T. Nakahata, F. Ishino, A. Ogura, and T. Shinohara. 2004. "Generation of pluripotent stem cells from neonatal mouse testis." *Cell* 119 (7):1001–12. doi: 10.1016/j.cell.2004.11.011.

Kandilci, A., E. Mientjes, and G. Grosveld. 2004. "Effects of SET and SET-CAN on the differentiation of the human promonocytic cell line U937." *Leukemia* 18 (2):337–40. doi: 10.1038/sj.leu.2403227.

Kauko, O., S. Y. Imanishi, E. Kulesskiy, L. Yetukuri, T. D. Laajala, M. Sharma, K. Pavic, A. Aakula, C. Rupp, M. Jumppanen, P. Haapaniemi, L. Ruan, B. Yadav, V. Suni, T. Varila, G. L. Corthals, J. Reimand, K. Wennerberg, T. Aittokallio, and J. Westermarck. 2020. "Phosphoproteome and drug-response effects mediated by the three protein phosphatase 2A inhibitor proteins CIP2A, SET, and PME-1." *J Biol Chem* 295 (13):4194–4211. doi: 10.1074/jbc.RA119.011265.

Kauko, O., C. M. O'Connor, E. Kulesskiy, J. Sangodkar, A. Aakula, S. Izadmehr, L. Yetukuri, B. Yadav, A. Padzik, T. D. Laajala, P. Haapaniemi, M. Momeny, T. Varila, M. Ohlmeyer, T. Aittokallio, K. Wennerberg, G. Narla, and J. Westermarck. 2018. "PP2A inhibition is a druggable MEK inhibitor resistance mechanism in KRAS-mutant lung cancer cells." *Sci Transl Med* 10 (450). doi: 10.1126/scitranslmed.aaq1093.

Kauko, O., and J. Westermarck. 2018. "Non-genomic mechanisms of protein phosphatase 2A (PP2A) regulation in cancer." *Int J Biochem Cell Biol* 96:157–164. doi: 10.1016/j.biocel.2018.01.005.

Kaur, A., O. V. Denisova, X. Qiao, M. Jumppanen, E. Peuhu, S. U. Ahmed, O. Raheem, H. Haapasalo, J. Eriksson, A. J. Chalmers, P. Laakkonen, and J. Westermarck. 2016. "PP2A Inhibitor PME-1

Drives Kinase Inhibitor Resistance in Glioma Cells." *Cancer Res* 76 (23):7001–7011. doi: 10.1158/0008-5472.CAN-16-1134.

Kaur, A., and J. Westermarck. 2016. "Regulation of protein phosphatase 2A (PP2A) tumor suppressor function by PME-1." *Biochem Soc Trans* 44 (6):1683–1693. doi: 10.1042/BST20160161.

Khanna, A., O. Kauko, C. Böckelman, A. Laine, I. Schreck, J. I. Partanen, A. Szwajda, S. Bormann, T. Bilgen, M. Helenius, Y. R. Pokharel, J. Pimanda, M. R. Russel, C. Haglund, K. A. Cole, J. Klefström, T. Aittokallio, C. Weiss, A. Ristimäki, T. Visakorpi, and J. Westermarck. 2013. "Chk1 targeting reactivates PP2A tumor suppressor activity in cancer cells." *Cancer Res* 73 (22):6757–69. doi: 10.1158/0008-5472.CAN-13-1002.

Khanna, A., J. Okkeri, T. Bilgen, T. Tiirikka, M. Vihinen, T. Visakorpi, and J. Westermarck. 2011. "ETS1 mediates MEK1/2-dependent overexpression of cancerous inhibitor of protein phosphatase 2A (CIP2A) in human cancer cells." *PLoS One* 6 (3):e17979. doi: 10.1371/journal.pone.0017979.

Khanna, A., and J. E. Pimanda. 2016. "Clinical significance of cancerous inhibitor of protein phosphatase 2A in human cancers." *Int J Cancer* 138 (3):525–32. doi: 10.1002/ijc.29431.

Khanna, A., J. E. Pimanda, and J. Westermarck. 2013. "Cancerous inhibitor of protein phosphatase 2A, an emerging human oncoprotein and a potential cancer therapy target." *Cancer Res* 73 (22):6548–53. doi: 10.1158/0008-5472.CAN-13-1994.

Khoury, G., R. Baliban, and C. Floudas. 2011. "Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database." *Scientific reports* 1:90.

Kim, J. Y., K. S. Lee, J. E. Seol, K. Yu, D. Chakravarti, and S. B. Seo. 2012. "Inhibition of p53 acetylation by INHAT subunit SET/TAF-Iβ represses p53 activity." *Nucleic Acids Res* 40 (1):75–87. doi: 10.1093/nar/gkr614.

Kim, T., M. S. Tyndel, H. J. Kim, J. S. Ahn, S. H. Choi, H. J. Park, Y. K. Kim, S. Y. Kim, J. H. Lipton, Z. Zhang, and D. D. Kim. 2017. "Spectrum of somatic mutation dynamics in chronic myeloid leukemia following tyrosine kinase inhibitor therapy." *Blood* 129 (1):38–47. doi: 10.1182/blood-2016-04-708560.

Ko, K., N. Tapia, G. Wu, J. B. Kim, M. J. Bravo, P. Sasse, T. Glaser, D. Ruau, D. W. Han, B. Greber, K. Hausdörfer, V. Sebastiano, M. Stehling, B. K. Fleischmann, O. Brüstle, M. Zenke, and H. R. Schöler. 2009. "Induction of pluripotency in adult unipotent germline stem cells." *Cell Stem Cell* 5 (1):87–96. doi: 10.1016/j.stem.2009.05.025.

Kreso, A., and J. E. Dick. 2014. "Evolution of the cancer stem cell model." *Cell Stem Cell* 14 (3):275–91. doi: 10.1016/j.stem.2014.02.006.

Kruse, T., S. P. Gnosa, I. Nasa, D. H. Garvanska, J. B. Hein, H. Nguyen, J. Samsøe-Petersen, B. Lopez-Mendez, E. P. T. Hertz, J. Schwarz, H. S. Pena, D. Nikodemus, M. Kveiborg, A. N. Kettenbach, and J. Nilsson. 2020. "Mechanisms of site-specific dephosphorylation and kinase opposition imposed by PP2A regulatory subunits." *EMBO J* 39 (13):e103695. doi: 10.15252/embj.2019103695.

Laine, A., H. Sihto, C. Come, M. T. Rosenfeldt, A. Zwolinska, M. Niemelä, A. Khanna, E. K. Chan, V. M. Kähäri, P. L. Kellokumpu-Lehtinen, O. J. Sansom, G. I. Evan, M. R. Junttila, K. M. Ryan, J. C. Marine, H. Joensuu, and J. Westermarck. 2013. "Senescence sensitivity of breast cancer cells is defined by positive feedback loop between CIP2A and E2F1." *Cancer Discov* 3 (2):182–97. doi: 10.1158/2159-8290.CD-12-0292.

Lansford, C., R. Grénman, H. Bier, K.D. Somers, S.-Y. Kim, T.L. Whiteside, G.L. Clayman, H.-J. Welkoborsky, and T.E. Carey. 1999. Head and Neck Cancers. Dordrecht: Kluwer Academic Press.

Lapidot, T., C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M. A. Caligiuri, and J. E. Dick. 1994. "A cell initiating human acute myeloid leukaemia after transplantation into SCID mice." *Nature* 367 (6464):645–8. doi: 10.1038/367645a0.

Laverdière, I., M. Boileau, A. L. Neumann, H. Frison, A. Mitchell, S. W. K. Ng, J. C. Y. Wang, M. D. Minden, and K. Eppert. 2018. "Leukemic stem cell signatures identify novel therapeutics targeting acute myeloid leukemia." *Blood Cancer J* 8 (6):52. doi: 10.1038/s41408-018-0087-2.

Lawrence, M. S., P. Stojanov, C. H. Mermel, J. T. Robinson, L. A. Garraway, T. R. Golub, M. Meyerson, S. B. Gabriel, E. S. Lander, and G. Getz. 2014. "Discovery and saturation analysis of cancer genes across 21 tumour types." *Nature* 505 (7484):495–501. doi: 10.1038/nature12912.

Lechner, M., G. M. Frampton, T. Fenton, A. Feber, G. Palmer, A. Jay, N. Pillay, M. Forster, M. T. Cronin, D. Lipson, V. A. Miller, T. A. Brennan, S. Henderson, F. Vaz, P. O'Flynn, N. Kalavrezos, R. Yelensky, S. Beck, P. J. Stephens, and C. Boshoff. 2013. "Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors." *Genome Med* 5 (5):49. doi: 10.1186/gm453.

Lee, J., and J. Stock. 1993. "Protein phosphatase 2A catalytic subunit is methyl-esterified at its carboxyl terminus by a novel methyltransferase." *J Biol Chem* 268 (26):19192-5.

Ley, T. J., C. Miller, L. Ding, B. J. Raphael, A. J. Mungall, A. Robertson, K. Hoadley, T. J. Triche, P. W. Laird, J. D. Baty, L. L. Fulton, R. Fulton, S. E. Heath, J. Kalicki-Veizer, C. Kandoth, J. M. Klco, D. C. Koboldt, K. L. Kanchi, S. Kulkarni, T. L. Lamprecht, D. E. Larson, L. Lin, C. Lu, M. D. McLellan, J. F. McMichael, J. Payton, H. Schmidt, D. H. Spencer, M. H. Tomasson, J. W. Wallis, L. D. Wartman, M. A. Watson, J. Welch, M. C. Wendl, A. Ally, M. Balasundaram, I. Birol, Y. Butterfield, R. Chiu, A. Chu, E. Chuah, H. J. Chun, R. Corbett, N. Dhalla, R. Guin, A. He, C. Hirst, M. Hirst, R. A. Holt, S. Jones, A. Karsan, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, K. Mungall, J. Parker, E. Pleasance, P. Plettner, J. Schein, D. Stoll, L. Swanson, A. Tam, N. Thiessen, R. Varhol, N. Wye, Y. Zhao, S. Gabriel, G. Getz, C. Sougnez, L. Zou, M. D. Leiserson, F. Vandin, H. T. Wu, F. Applebaum, S. B. Baylin, R. Akbani, B. M. Broom, K. Chen, T. C. Motter, K. Nguyen, J. N. Weinstein, N. Zhang, M. L. Ferguson, C. Adams, A. Black, J. Bowen, J. Gastier-Foster, T. Grossman, T. Lichtenberg, L. Wise, T. Davidsen, J. A. Demchok, K. R. Shaw, M. Sheth, H. J. Sofia, L. Yang, J. R. Downing, G. Eley, and Cancer Genome Atlas Research Network. 2013. "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." *N Engl J Med* 368 (22):2059–74. doi: 10.1056/NEJMoa1301689.

Li, M., A. Makkinje, and Z. Damuni. 1996. "The myeloid leukemia-associated protein SET is a potent inhibitor of protein phosphatase 2A." *J Biol Chem* 271 (19):11059–62. doi: 10.1074/jbc.271.19.11059.

Li, X., M. Wilmanns, J. Thornton, and M. Köhn. 2013. "Elucidating human phosphatase-substrate networks." *Sci Signal* 6 (275):rs10. doi: 10.1126/scisignal.2003203.

Lindsley, R. C., B. G. Mar, E. Mazzola, P. V. Grauman, S. Shareef, S. L. Allen, A. Pigneux, M. Wetzler, R. K. Stuart, H. P. Erba, L. E. Damon, B. L. Powell, N. Lindeman, D. P. Steensma, M. Wadleigh, D. J. DeAngelo, D. Neuberg, R. M. Stone, and B. L. Ebert. 2015. "Acute myeloid leukemia ontogeny is defined by distinct somatic mutations." *Blood* 125 (9):1367–76. doi: 10.1182/blood-2014-11-610543.

Liu, J., B. Huang, Y. Xiao, H. M. Xiong, J. Li, D. Q. Feng, X. M. Chen, H. B. Zhang, and X. Z. Wang. 2012. "Aberrant expression of splicing factors in newly diagnosed acute myeloid leukemia." *Onkologie* 35 (6):335–40. doi: 10.1159/000338941.

Liu, S., A. E. Marneth, G. Alexe, S. R. Walker, H. I. Gandler, D. Q. Ye, K. Labella, R. Mathur, P. A. Toniolo, M. Tillgren, P. C. Gokhale, D. Barbie, A. Mullally, K. Stegmaier, and D. A. Frank. 2018. "The kinases IKBKE and TBK1 regulate MYC-dependent survival pathways through YB-1 in AML and are targets for therapy." *Blood Adv* 2 (23):3428–3442. doi: 10.1182/bloodadvances.2018016733.

Longin, S., J. Jordens, E. Martens, I. Stevens, V. Janssens, E. Rondelez, I. De Baere, R. Derua, E. Waelkens, J. Goris, and C. Van Hoof. 2004. "An inactive protein phosphatase 2A population is associated with methylesterase and can be re-activated by the phosphotyrosyl phosphatase activator." *Biochem J* 380 (Pt 1):111–9. doi: 10.1042/BJ20031643.

Longin, S., K. Zwaenepoel, J. V. Louis, S. Dilworth, J. Goris, and V. Janssens. 2007. "Selection of protein phosphatase 2A regulatory subunits is mediated by the C terminus of the catalytic Subunit." *J Biol Chem* 282 (37):26971–80. doi: 10.1074/jbc.M704059200.

Lucas, C. M., R. J. Harris, A. Giannoudis, M. Copland, J. R. Slupsky, and R. E. Clark. 2011. "Cancerous inhibitor of PP2A (CIP2A) at diagnosis of chronic myeloid leukemia is a critical determinant of disease progression." *Blood* 117 (24):6660–8. doi: 10.1182/blood-2010-08-304477.

Lucas, C. M., R. J. Harris, A. K. Holcroft, L. J. Scott, N. Carmell, E. McDonald, F. Polydoros, and R. E. Clark. 2015. "Second generation tyrosine kinase inhibitors prevent disease progression in high-risk (high CIP2A) chronic myeloid leukaemia patients." *Leukemia* 29 (7):1514–23. doi: 10.1038/leu.2015.71.

Lucas, C. M., L. J. Scott, N. Carmell, A. K. Holcroft, R. K. Hills, A. K. Burnett, and R. E. Clark. 2018. "CIP2A- and SETBP1-mediated PP2A inhibition reveals AKT S473 phosphorylation to be a new biomarker in AML." *Blood Adv* 2 (9):964–968. doi: 10.1182/bloodadvances.2017013615.

Lugo, T. G., A. M. Pendergast, A. J. Muller, and O. N. Witte. 1990. "Tyrosine kinase activity and transformation potency of bcr-abl oncogene products." *Science* 247 (4946):1079–82. doi: 10.1126/science.2408149.

Lui, V. W., M. L. Hedberg, H. Li, B. S. Vangara, K. Pendleton, Y. Zeng, Y. Lu, Q. Zhang, Y. Du, B. R. Gilbert, M. Freilino, S. Sauerwein, N. D. Peyser, D. Xiao, B. Diergaarde, L. Wang, S. Chiosea, R. Seethala, J. T. Johnson, S. Kim, U. Duvvuri, R. L. Ferris, M. Romkes, T. Nukui, P. Kwok-Shing Ng, L. A. Garraway, P. S. Hammerman, G. B. Mills, and J. R. Grandis. 2013. "Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers." *Cancer Discov* 3 (7):761–9. doi: 10.1158/2159-8290.CD-13-0103.

Lü, M., K. Ding, G. Zhang, M. Yin, G. Yao, H. Tian, J. Lian, L. Liu, M. Liang, T. Zhu, and F. Sun. 2015. "MicroRNA-320a sensitizes tamoxifen-resistant breast cancer cells to tamoxifen by targeting ARPP-19 and ERRγ." *Sci Rep* 5:8735. doi: 10.1038/srep08735.

Ma, Y., Y. Liu, Y. S. Pu, M. L. Cui, Z. J. Mao, Z. Z. Li, L. He, M. Wu, and J. H. Wang. 2020. "LncRNA IGFL2-AS1 functions as a ceRNA in regulating ARPP19 through competitive binding to miR-802 in gastric cancer." *Mol Carcinog* 59 (3):311–322. doi: 10.1002/mc.23155.

Malta, T. M., A. Sokolov, A. J. Gentles, T. Burzykowski, L. Poisson, J. N. Weinstein, B. Kamińska, J. Huelsken, L. Omberg, O. Gevaert, A. Colaprico, P. Czerwińska, S. Mazurek, L. Mishra, H. Heyn, A. Krasnitz, A. K. Godwin, A. J. Lazar, J. M. Stuart, K. A. Hoadley, P. W. Laird, H. Noushmehr, M. Wiznerowicz, and Cancer Genome Atlas Research Network. 2018. "Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation." *Cell* 173 (2):338–354.e15. doi: 10.1016/j.cell.2018.03.034.

Manchado, E., M. Guillamot, G. de Cárcer, M. Eguren, M. Trickey, I. García-Higuera, S. Moreno, H. Yamano, M. Cañamero, and M. Malumbres. 2010. "Targeting mitotic exit leads to tumor regression in vivo: Modulation by Cdk1, Mastl, and the PP2A/B55α,δ phosphatase." *Cancer Cell* 18 (6):641–54. doi: 10.1016/j.ccr.2010.10.028.

Manning, G., D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. 2002. "The protein kinase complement of the human genome." *Science* 298 (5600):1912–34. doi: 10.1126/science.1075762.

Mansour, W. Y., N. V. Bogdanova, U. Kasten-Pisula, T. Rieckmann, S. Köcher, K. Borgmann, M. Baumann, M. Krause, C. Petersen, H. Hu, R. A. Gatti, E. Dikomey, T. Dörk, and J. Dahm-Daphi. 2013. "Aberrant overexpression of miR-421 downregulates ATM and leads to a pronounced DSB

repair defect and clinical hypersensitivity in SKX squamous cell carcinoma." *Radiother Oncol* 106 (1):147–54. doi: 10.1016/j.radonc.2012.10.020.

Mariotto, A. B., L. Enewold, J. Zhao, C. A. Zeruto, and K. R. Yabroff. 2020. "Medical Care Costs Associated with Cancer Survivorship in the United States." *Cancer Epidemiol Biomarkers Prev* 29 (7):1304–1312. doi: 10.1158/1055-9965.EPI-19-1534.

Martínez-Jiménez, F., F. Muiños, I. Sentís, J. Deu-Pons, I. Reyes-Salazar, C. Arnedo-Pac, L. Mularoni, O. Pich, J. Bonet, H. Kranas, A. Gonzalez-Perez, and N. Lopez-Bigas. 2020. "A compendium of mutational cancer driver genes." *Nat Rev Cancer* 20 (10):555–572. doi: 10.1038/s41568-020-0290-x.

Matthews, H. R. 1995. "Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: a possible regulator of the mitogen-activated protein kinase cascade." In *Pharmacol Ther*, 323–50. England.

McConnell, J. L., R. J. Gomez, L. R. McCorvey, B. K. Law, and B. E. Wadzinski. 2007. "Identification of a PP2A-interacting protein that functions as a negative regulator of phosphatase activity in the ATM/ATR signaling pathway." *Oncogene* 26 (41):6021–30. doi: 10.1038/sj.onc.1210406.

McIntosh, B. E., M. E. Brown, B. M. Duffin, J. P. Maufort, D. T. Vereide, I. I. Slukvin, and J. A. Thomson. 2015. "Nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) (NBSGW) mice support multilineage engraftment of human hematopoietic cells." *Stem Cell Reports* 4 (2):171–80. doi: 10.1016/j.stemcr.2014.12.005.

Meeusen, B., and V. Janssens. 2018. "Tumor suppressive protein phosphatases in human cancer: Emerging targets for therapeutic intervention and tumor stratification." *Int J Biochem Cell Biol* 96:98–134. doi: 10.1016/j.biocel.2017.10.002.

Merlos-Suárez, A., F. M. Barriga, P. Jung, M. Iglesias, M. V. Céspedes, D. Rossell, M. Sevillano, X. Hernando-Momblona, V. da Silva-Diz, P. Muñoz, H. Clevers, E. Sancho, R. Mangues, and E. Batlle. 2011. "The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse." *Cell Stem Cell* 8 (5):511–24. doi: 10.1016/j.stem.2011.02.020.

Miranda, A., P. T. Hamilton, A. W. Zhang, S. Pattnaik, E. Becht, A. Mezheyeuski, J. Bruun, P. Micke, A. de Reynies, and B. H. Nelson. 2019. "Cancer stemness, intratumoral heterogeneity, and immune response across cancers." *Proc Natl Acad Sci U S A* 116 (18):9020–9029. doi: 10.1073/pnas.1818210116.

Mochida, S., S. L. Maslen, M. Skehel, and T. Hunt. 2010. "Greatwall phosphorylates an inhibitor of protein phosphatase 2A that is essential for mitosis." *Science* 330 (6011):1670–3. doi: 10.1126/science.1195689.

Moeller, B. J., J. S. Yordy, M. D. Williams, U. Giri, U. Raju, D. P. Molkentine, L. A. Byers, J. V. Heymach, M. D. Story, J. J. Lee, E. M. Sturgis, R. S. Weber, A. S. Garden, K. K. Ang, and D. L. Schwartz. 2011. "DNA repair biomarker profiling of head and neck cancer: Ku80 expression predicts locoregional failure and death following radiotherapy." *Clin Cancer Res* 17 (7):2035–43. doi: 10.1158/1078-0432.CCR-10-2641.

Mosca, L., P. Musto, K. Todoerti, M. Barbieri, L. Agnelli, S. Fabris, G. Tuana, M. Lionetti, E. Bonaparte, S. M. Sirchia, V. Grieco, G. Bianchino, F. D'Auria, T. Statuto, C. Mazzoccoli, L. De Luca, M. T. Petrucci, F. Morabito, M. Offidani, F. Di Raimondo, A. Falcone, T. Caravita, P. Omedè, M. Boccadoro, A. Palumbo, and A. Neri. 2013. "Genome-wide analysis of primary plasma cell leukemia identifies recurrent imbalances associated with changes in transcriptional profiles." *Am J Hematol* 88 (1):16–23. doi: 10.1002/ajh.23339.

Mrózek, K., K. Heinonen, A. de la Chapelle, and C. D. Bloomfield. 1997. "Clinical significance of cytogenetics in acute myeloid leukemia." *Semin Oncol* 24 (1):17–31.

Murakami, S., W. Ninomiya, E. Sakamoto, T. Shibata, H. Akiyama, and F. Tashiro. 2015. "SRY and OCT4 Are Required for the Acquisition of Cancer Stem Cell-Like Properties and Are Potential Differentiation Therapy Targets." *Stem Cells* 33 (9):2652–63. doi: 10.1002/stem.2059.

Musante, V., L. Li, J. Kanyo, T. T. Lam, C. M. Colangelo, S. K. Cheng, A. H. Brody, P. Greengard, N. Le Novère, and A. C. Nairn. 2017. "Reciprocal regulation of ARPP-16 by PKA and MAST3 kinases provides a cAMP-regulated switch in protein phosphatase 2A inhibition." *Elife* 6. doi: 10.7554/eLife.24998.

Myant, K., X. Qiao, T. Halonen, C. Come, A. Laine, M. Janghorban, J. I. Partanen, J. Cassidy, E. L. Ogg, P. Cammareri, T. Laiterä, J. Okkeri, J. Klefström, R. C. Sears, O. J. Sansom, and J. Westermarck. 2015. "Serine 62-Phosphorylated MYC Associates with Nuclear Lamins and Its Regulation by CIP2A Is Essential for Regenerative Proliferation." *Cell Rep* 12 (6):1019–31. doi: 10.1016/j.celrep.2015.07.003.

Mäkelä, E., E. Löyttyniemi, U. Salmenniemi, O. Kauko, T. Varila, V. Kairisto, M. Itälä-Remes, and J. Westermarck. 2019. "Arpp19 Promotes Myc and Cip2a Expression and Associates with Patient Relapse in Acute Myeloid Leukemia." *Cancers (Basel)* 11 (11). doi: 10.3390/cancers11111774.

Nakashima, A., K. Tanimura-Ito, N. Oshiro, S. Eguchi, T. Miyamoto, A. Momonami, S. Kamada, K. Yonezawa, and U. Kikkawa. 2013. "A positive role of mammalian Tip41-like protein, TIPRL, in the amino-acid dependent mTORC1-signaling pathway through interaction with PP2A." *FEBS Lett* 587 (18):2924–9. doi: 10.1016/j.febslet.2013.07.027.

Neviani, P., R. Santhanam, J. J. Oaks, A. M. Eiring, M. Notari, B. W. Blaser, S. Liu, R. Trotta, N. Muthusamy, C. Gambacorti-Passerini, B. J. Druker, J. Cortes, G. Marcucci, C. S. Chen, N. M. Verrills, D. C. Roy, M. A. Caligiuri, C. D. Bloomfield, J. C. Byrd, and D. Perrotti. 2007. "FTY720, a new alternative for treating blast crisis chronic myelogenous leukemia and Philadelphia chromosome-positive acute lymphocytic leukemia." *J Clin Invest* 117 (9):2408–21. doi: 10.1172/JCI31095.

Neviani, P., R. Santhanam, R. Trotta, M. Notari, B. W. Blaser, S. Liu, H. Mao, J. S. Chang, A. Galietta, A. Uttam, D. C. Roy, M. Valtieri, R. Bruner-Klisovic, M. A. Caligiuri, C. D. Bloomfield, G. Marcucci, and D. Perrotti. 2005. "The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein." *Cancer Cell* 8 (5):355–68. doi: 10.1016/j.ccr.2005.10.015.

Ng, S. W., A. Mitchell, J. A. Kennedy, W. C. Chen, J. McLeod, N. Ibrahimova, A. Arruda, A. Popescu, V. Gupta, A. D. Schimmer, A. C. Schuh, K. W. Yee, L. Bullinger, T. Herold, D. Görlich, T. Büchner, W. Hiddemann, W. E. Berdel, B. Wörmann, M. Cheok, C. Preudhomme, H. Dombret, K. Metzeler, C. Buske, B. Löwenberg, P. J. Valk, P. W. Zandstra, M. D. Minden, J. E. Dick, and J. C. Wang. 2016. "A 17-gene stemness score for rapid determination of risk in acute leukaemia." *Nature* 540 (7633):433–437. doi: 10.1038/nature20598.

Nicolini, F. E., S. Dulucq, L. Boureau, P. Cony-Makhoul, A. Charbonnier, M. Escoffre-Barbe, F. Rigal-Huguet, V. Coiteux, B. Varet, V. Dubruille, P. Lenain, P. Rousselot, D. Rea, A. Guerci-Bresler, L. Legros, J. Liu, M. Gardembas, J. C. Ianotto, P. Turlure, H. Johnson-Ansah, J. Martiniuc, H. Jardel, B. Joly, P. Zunic, T. Henni, B. Villemagne, M. G. Berger, E. Cayssials, F. Guilhot, F. Larosa, J. Guilhot, G. Etienne, and F. X. Mahon. 2019. "Evaluation of Residual Disease and TKI Duration Are Critical Predictive Factors for Molecular Recurrence after Stopping Imatinib First-line in Chronic Phase CML Patients." *Clin Cancer Res* 25 (22):6606–6613. doi: 10.1158/1078-0432.CCR-18-3373.

Niwa, H., J. Miyazaki, and A. G. Smith. 2000. "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." *Nat Genet* 24 (4):372–6. doi: 10.1038/74199.

Nunbhakdi-Craig, V., S. Schuechner, J. M. Sontag, L. Montgomery, D. C. Pallas, C. Juno, I. Mudrak, E. Ogris, and E. Sontag. 2007. "Expression of protein phosphatase 2A mutants and silencing of the regulatory B alpha subunit induce a selective loss of acetylated and detyrosinated microtubules." *J Neurochem* 101 (4):959–71. doi: 10.1111/j.1471-4159.2007.04503.x.

Ogris, E., X. Du, K. C. Nelson, E. K. Mak, X. X. Yu, W. S. Lane, and D. C. Pallas. 1999. "A protein phosphatase methylesterase (PME-1) is one of several novel proteins stably associating with two inactive mutants of protein phosphatase 2A." *J Biol Chem* 274 (20):14382–91. doi: 10.1074/jbc.274.20.14382.

O'Brien, S., L. Cork, V. Bandiera, R. Bescoby, L. Foroni, L. Alaily, W. Osborne, H. Bell-Gorrod, N. Latimer, JF. Apperley, C. Hedgley, R. Szydlo, J. Byrne, C. Pocock, B. Ramsahoye, T. Zwingers, J. Wason, M. Copland, and RE. Clark. 2018. SPIRIT 2: Final 5 year analysis of the UK National Cancer Research Institute randomized study comparing imatinib with dasatinib in patients with newly diagnosed chronic phase chronic myeloid leukemia.: Blood.

Padula, W. V., R. A. Larson, S. B. Dusetzina, J. F. Apperley, R. Hehlmann, M. Baccarani, E. Eigendorff, J. Guilhot, F. Guilhot, F. X. Mahon, G. Martinelli, J. Mayer, M. C. Müller, D. Niederwieser, S. Saussele, C. A. Schiffer, R. T. Silver, B. Simonsson, and R. M. Conti. 2016. "Cost-effectiveness of Tyrosine Kinase Inhibitor Treatment Strategies for Chronic Myeloid Leukemia in Chronic Phase After Generic Entry of Imatinib in the United States." *J Natl Cancer Inst* 108 (7). doi: 10.1093/jnci/djw003.

Pallas, D. C., L. K. Shahrik, B. L. Martin, S. Jaspers, T. B. Miller, D. L. Brautigan, and T. M. Roberts. 1990. "Polyoma small and middle T antigens and SV40 small t antigen form stable complexes with protein phosphatase 2A." *Cell* 60 (1):167–76. doi: 10.1016/0092-8674(90)90726-u.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe. 2008. "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." *Nat Genet* 40 (12):1413–5. doi: 10.1038/ng.259.

Papaemmanuil, E., M. Gerstung, L. Bullinger, V. I. Gaidzik, P. Paschka, N. D. Roberts, N. E. Potter, M. Heuser, F. Thol, N. Bolli, G. Gundem, P. Van Loo, I. Martincorena, P. Ganly, L. Mudie, S. McLaren, S. O'Meara, K. Raine, D. R. Jones, J. W. Teague, A. P. Butler, M. F. Greaves, A. Ganser, K. Döhner, R. F. Schlenk, H. Döhner, and P. J. Campbell. 2016. "Genomic Classification and Prognosis in Acute Myeloid Leukemia." *N Engl J Med* 374 (23):2209–2221. doi: 10.1056/NEJMoa1516192.

Parkinson, D. R., R. T. McCormack, S. M. Keating, S. I. Gutman, S. R. Hamilton, E. A. Mansfield, M. A. Piper, P. Deverka, F. W. Frueh, J. M. Jessup, L. M. McShane, S. R. Tunis, C. C. Sigman, and G. J. Kelloff. 2014. "Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer." *Clin Cancer Res* 20 (6):1428-44. doi: 10.1158/1078-0432.CCR-13-2961.

Paz, I., M. Akerman, I. Dror, I. Kosti, and Y. Mandel-Gutfreund. 2010. "SFmap: a web server for motif analysis and prediction of splicing factor binding sites." *Nucleic Acids Res* 38 (Web Server issue):W281-5. doi: 10.1093/nar/gkq444.

Pece, S., D. Disalvatore, D. Tosoni, M. Vecchi, S. Confalonieri, G. Bertalot, G. Viale, M. Colleoni, P. Veronesi, V. Galimberti, and P. P. Di Fiore. 2019. "Identification and clinical validation of a multigene assay that interrogates the biology of cancer stem cells and predicts metastasis in breast cancer: A retrospective consecutive study." *EBioMedicine* 42:352–362. doi: 10.1016/j.ebiom.2019.02.036.

Peitzsch, C., J. Nathansen, S. I. Schniewind, F. Schwarz, and A. Dubrovska. 2019. "Cancer Stem Cells in Head and Neck Squamous Cell Carcinoma: Identification, Characterization and Clinical Implications." *Cancers (Basel)* 11 (5). doi: 10.3390/cancers11050616.

Pekkola-Heino, K., J. Kulmala, P. Klemi, T. Lakkala, K. Aitasalo, H. Joensuu, and R. Grenman. 1991. "Effects of radiation fractionation on four squamous cell carcinoma lines with dissimilar inherent radiation sensitivity." *J Cancer Res Clin Oncol* 117 (6):597–602. doi: 10.1007/BF01613295.

Pekkola-Heino, K., K. Servomaa, A. Kiuru, and R. Grenman. 1998. "Sublethal damage repair capacity in carcinoma cell lines with p53 mutations." *Head Neck* 20 (4):298–303. doi: 10.1002/(sici)1097-0347(199807)20:4<298::aid-hed3>3.0.co;2-u.

Perrotti, D., and P. Neviani. 2013. "Protein phosphatase 2A: a target for anticancer therapy." *Lancet Oncol* 14 (6):e229–38. doi: 10.1016/S1470-2045(12)70558-2.

Petersdorf, S. H., K. J. Kopecky, M. Slovak, C. Willman, T. Nevill, J. Brandwein, R. A. Larson, H. P. Erba, P. J. Stiff, R. K. Stuart, R. B. Walter, M. S. Tallman, L. Stenke, and F. R. Appelbaum. 2013. "A phase 3 study of gemtuzumab ozogamicin during induction and postconsolidation therapy in younger patients with acute myeloid leukemia." *Blood* 121 (24):4854–60. doi: 10.1182/blood-2013-01-466706.

Piva, F., M. Giulietti, A. B. Burini, and G. Principato. 2012. "SpliceAid 2: a database of human splicing factors expression data and RNA target motifs." *Hum Mutat* 33 (1):81–5. doi: 10.1002/humu.21609.

Plaks, V., N. Kong, and Z. Werb. 2015. "The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells?" *Cell Stem Cell* 16 (3):225–38. doi: 10.1016/j.stem.2015.02.015.

Ponti, D., A. Costa, N. Zaffaroni, G. Pratesi, G. Petrangolini, D. Coradini, S. Pilotti, M. A. Pierotti, and M. G. Daidone. 2005. "Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties." *Cancer Res* 65 (13):5506–11. doi: 10.1158/0008-5472.CAN-05-0626.

Puustinen, P., M. R. Junttila, S. Vanhatupa, A. A. Sablina, M. E. Hector, K. Teittinen, O. Raheem, K. Ketola, S. Lin, J. Kast, H. Haapasalo, W. C. Hahn, and J. Westermarck. 2009. "PME-1 protects extracellular signal-regulated kinase pathway activity from protein phosphatase 2A-mediated inactivation in human malignant glioma." *Cancer Res* 69 (7):2870–7. doi: 10.1158/0008-5472.CAN-08-2760.

Quintana, E., M. Shackleton, M. S. Sabel, D. R. Fullen, T. M. Johnson, and S. J. Morrison. 2008. "Efficient tumour formation by single human melanoma cells." *Nature* 456 (7222):593–8. doi: 10.1038/nature07567.

Radich, J. P., M. Deininger, C. N. Abboud, J. K. Altman, E. Berman, R. Bhatia, B. Bhatnagar, P. Curtin, D. J. DeAngelo, J. Gotlib, G. Hobbs, M. Jagasia, H. M. Kantarjian, L. Maness, L. Metheny, J. O. Moore, A. Pallera, P. Pancari, M. Patnaik, E. Purev, M. G. Rose, N. P. Shah, B. D. Smith, D. S. Snyder, K. L. Sweet, M. Talpaz, J. Thompson, D. T. Yang, K. M. Gregory, and H. Sundar. 2018. "Chronic Myeloid Leukemia, Version 1.2019, NCCN Clinical Practice Guidelines in Oncology." *J Natl Compr Canc Netw* 16 (9):1108–1135. doi: 10.6004/jnccn.2018.0071.

Rangarajan, A., S. J. Hong, A. Gifford, and R. A. Weinberg. 2004. "Species- and cell type-specific requirements for cellular transformation." *Cancer Cell* 6 (2):171–83. doi: 10.1016/j.ccr.2004.07.009.

Robb, M. A., P. M. McInnes, and R. M. Califf. 2016. "Biomarkers and Surrogate Endpoints: Developing Common Terminology and Definitions." *JAMA* 315 (11):1107–8. doi: 10.1001/jama.2016.2240.

Routila, J., T. Bilgen, O. Saramäki, R. Grénman, T. Visakorpi, J. Westermarck, and S. Ventelä. 2016. "Copy number increase of oncoprotein CIP2A is associated with poor patient survival in human head and neck squamous cell carcinoma." *J Oral Pathol Med* 45 (5):329–37. doi: 10.1111/jop.12372.

Sablina, A. A., M. Hector, N. Colpaert, and W. C. Hahn. 2010. "Identification of PP2A complexes and pathways involved in cell transformation." *Cancer Res* 70 (24):10474–84. doi: 10.1158/0008-5472.CAN-10-2855.

Saito, S., M. Miyaji-Yamaguchi, T. Shimoyama, and K. Nagata. 1999. "Functional domains of template-activating factor-I as a protein phosphatase 2A inhibitor." *Biochem Biophys Res Commun* 259 (2):471–5. doi: 10.1006/bbrc.1999.0790.

Sangodkar, J., C. C. Farrington, K. McClinch, M. D. Galsky, D. B. Kastrinsky, and G. Narla. 2016. "All roads lead to PP2A: exploiting the therapeutic potential of this phosphatase." *FEBS J* 283 (6):1004–24. doi: 10.1111/febs.13573.

Sawyers, C. L., and L. J. van 't Veer. 2014. "Reliable and effective diagnostics are keys to accelerating personalized cancer medicine and transforming cancer care: a policy statement from the american association for cancer research." *Clin Cancer Res* 20 (19):4978–81. doi: 10.1158/1078-0432.CCR-14-2295.

Schuurhuis, G. J., M. Heuser, S. Freeman, M. C. Béné, F. Buccisano, J. Cloos, D. Grimwade, T. Haferlach, R. K. Hills, C. S. Hourigan, J. L. Jorgensen, W. Kern, F. Lacombe, L. Maurillo, C. Preudhomme, B. A. van der Reijden, C. Thiede, A. Venditti, P. Vyas, B. L. Wood, R. B. Walter, K. Döhner, G. J. Roboz, and G. J. Ossenkoppele. 2018. "Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party." *Blood* 131 (12):1275–1291. doi: 10.1182/blood-2017-09-801498.

Scott, M. T., K. Korfi, P. Saffrey, L. E. Hopcroft, R. Kinstrie, F. Pellicano, C. Guenther, P. Gallipoli, M. Cruz, K. Dunn, H. G. Jorgensen, J. E. Cassels, A. Hamilton, A. Crossan, A. Sinclair, T. L. Holyoake, and D. Vetrie. 2016. "Epigenetic Reprogramming Sensitizes CML Stem Cells to Combined EZH2 and Tyrosine Kinase Inhibition." *Cancer Discov* 6 (11):1248–1257. doi: 10.1158/2159-8290.CD-16-0263.

Sents, W., E. Ivanova, C. Lambrecht, D. Haesen, and V. Janssens. 2013. "The biogenesis of active protein phosphatase 2A holoenzymes: a tightly regulated process creating phosphatase specificity." *FEBS J* 280 (2):644–61. doi: 10.1111/j.1742-4658.2012.08579.x.

Seo, S. B., P. McNamara, S. Heo, A. Turner, W. S. Lane, and D. Chakravarti. 2001. "Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein." *Cell* 104 (1):119–30. doi: 10.1016/s0092-8674(01)00196-9.

Shao, M., T. Bi, W. Ding, C. Yu, C. Jiang, H. Yang, X. Sun, and M. Yang. 2018. "OCT4 Potentiates Radio-Resistance and Migration Activity of Rectal Cancer Cells by Improving Epithelial-Mesenchymal Transition in a ZEB1 Dependent Manner." *Biomed Res Int* 2018:3424956. doi: 10.1155/2018/3424956.

Shen, L., X. Huang, X. Xie, J. Su, J. Yuan, and X. Chen. 2014. "High Expression of SOX2 and OCT4 Indicates Radiation Resistance and an Independent Negative Prognosis in Cervical Squamous Cell Carcinoma." *J Histochem Cytochem* 62 (7):499–509. doi: 10.1369/0022155414532654.

Shi, Y. 2009. "Serine/threonine phosphatases: mechanism through structure." *Cell* 139 (3):468–84. doi: 10.1016/j.cell.2009.10.006.

Shih, Y. T., J. E. Cortes, and H. M. Kantarjian. 2019. "Treatment value of second-generation BCR-ABL1 tyrosine kinase inhibitors compared with imatinib to achieve treatment-free remission in patients with chronic myeloid leukaemia: a modelling study." *Lancet Haematol* 6 (8):e398–e408. doi: 10.1016/S2352-3026(19)30087-0.

Shimokawa, M., Y. Ohta, S. Nishikori, M. Matano, A. Takano, M. Fujii, S. Date, S. Sugimoto, T. Kanai, and T. Sato. 2017. "Visualization and targeting of LGR5." *Nature* 545 (7653):187–192. doi: 10.1038/nature22081.

Shin, K. S., E. Y. Shin, S. C. Bae, S. R. Kim, G. B. Jeong, S. J. Kwak, B. J. Ballermann, and E. G. Kim. 1999. "Expression of SET is modulated as a function of cell proliferation." *J Cell Biochem* 74 (1):119–26.

Shlush, L. I., A. Mitchell, L. Heisler, S. Abelson, S. W. K. Ng, A. Trotman-Grant, J. J. F. Medeiros, A. Rao-Bhatia, I. Jaciw-Zurakowsky, R. Marke, J. L. McLeod, M. Doedens, G. Bader, V. Voisin, C. Xu, J. D. McPherson, T. J. Hudson, J. C. Y. Wang, M. D. Minden, and J. E. Dick. 2017. "Tracing the origins of relapse in acute myeloid leukaemia to stem cells." *Nature* 547 (7661):104–108. doi: 10.1038/nature22993.

Shlush, L. I., S. Zandi, A. Mitchell, W. C. Chen, J. M. Brandwein, V. Gupta, J. A. Kennedy, A. D. Schimmer, A. C. Schuh, K. W. Yee, J. L. McLeod, M. Doedens, J. J. Medeiros, R. Marke, H. J. Kim, K. Lee, J. D. McPherson, T. J. Hudson, A. M. Brown, F. Yousif, Q. M. Trinh, L. D. Stein, M. D. Minden, J. C. Wang, J. E. Dick, and HALT Pan-Leukemia Gene Panel Consortium. 2014. "Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia." *Nature* 506 (7488):328–33. doi: 10.1038/nature13038.

Shouse, G., R. de Necochea-Campion, S. Mirshahidi, X. Liu, and C. S. Chen. 2016. "Novel B55α-PP2A mutations in AML promote AKT T308 phosphorylation and sensitivity to AKT inhibitor-induced growth arrest." *Oncotarget* 7 (38):61081–61092. doi: 10.18632/oncotarget.11209.

Shtrichman, R., R. Sharf, H. Barr, T. Dobner, and T. Kleinberger. 1999. "Induction of apoptosis by adenovirus E4orf4 protein is specific to transformed cells and requires an interaction with protein phosphatase 2A." *Proc Natl Acad Sci U S A* 96 (18):10080–5. doi: 10.1073/pnas.96.18.10080.

Siegel, R. L., K. D. Miller, and A. Jemal. 2020. "Cancer statistics, 2020." *CA Cancer J Clin* 70 (1):7–30. doi: 10.3322/caac.21590.

Smith, B. A., N. G. Balanis, A. Nanjundiah, K. M. Sheu, B. L. Tsai, Q. Zhang, J. W. Park, M. Thompson, J. Huang, O. N. Witte, and T. G. Graeber. 2018. "A Human Adult Stem Cell Signature Marks Aggressive Variants across Epithelial Cancers." *Cell Rep* 24 (12):3353–3366.e5. doi: 10.1016/j.celrep.2018.08.062.

Song, H., J. Pan, Y. Liu, H. Wen, L. Wang, J. Cui, B. Hu, Z. Yao, and G. Ji. 2014. "Increased ARPP-19 expression is associated with hepatocellular carcinoma." *Int J Mol Sci* 16 (1):178–92. doi: 10.3390/ijms16010178.

Song, I. S., S. Y. Jun, H. J. Na, H. T. Kim, S. Y. Jung, G. H. Ha, Y. H. Park, L. Z. Long, D. Y. Yu, J. M. Kim, J. H. Kim, J. H. Ko, C. H. Kim, and N. S. Kim. 2012. "Inhibition of MKK7-JNK by the TOR signaling pathway regulator-like protein contributes to resistance of HCC cells to TRAIL-induced apoptosis." *Gastroenterology* 143 (5):1341–1351. doi: 10.1053/j.gastro.2012.07.103.

Soo Hoo, L., J. Y. Zhang, and E. K. Chan. 2002. "Cloning and characterization of a novel 90 kDa 'companion' auto-antigen of p62 overexpressed in cancer." *Oncogene* 21 (32):5006–15. doi: 10.1038/sj.onc.1205625.

Soofiyani, S. R., M. S. Hejazi, and B. Baradaran. 2017. "The role of CIP2A in cancer: A review and update." *Biomed Pharmacother* 96:626–633. doi: 10.1016/j.biopha.2017.08.146.

Soverini, S., C. de Benedittis, M. Mancini, and G. Martinelli. 2015. "Mutations in the BCR-ABL1 Kinase Domain and Elsewhere in Chronic Myeloid Leukemia." *Clin Lymphoma Myeloma Leuk* 15 Suppl:S120–8. doi: 10.1016/j.clml.2015.02.035.

Stanevich, V., L. Jiang, K. A. Satyshur, Y. Li, P. D. Jeffrey, Z. Li, P. Menden, M. F. Semmelhack, and Y. Xing. 2011. "The structural basis for tight control of PP2A methylation and function by LCMT-1." *Mol Cell* 41 (3):331–42. doi: 10.1016/j.molcel.2010.12.030.

Steegmann, J. L., M. Baccarani, M. Breccia, L. F. Casado, V. García-Gutiérrez, A. Hochhaus, D. W. Kim, T. D. Kim, H. J. Khoury, P. Le Coutre, J. Mayer, D. Milojkovic, K. Porkka, D. Rea, G. Rosti, S. Saussele, R. Hehlmann, and R. E. Clark. 2016. "European LeukemiaNet recommendations for

the management and avoidance of adverse events of treatment in chronic myeloid leukaemia." *Leukemia* 30 (8):1648–71. doi: 10.1038/leu.2016.104.

Stransky, N., A. M. Egloff, A. D. Tward, A. D. Kostic, K. Cibulskis, A. Sivachenko, G. V. Kryukov, M. S. Lawrence, C. Sougnez, A. McKenna, E. Shefler, A. H. Ramos, P. Stojanov, S. L. Carter, D. Voet, M. L. Cortés, D. Auclair, M. F. Berger, G. Saksena, C. Guiducci, R. C. Onofrio, M. Parkin, M. Romkes, J. L. Weissfeld, R. R. Seethala, L. Wang, C. Rangel-Escareño, J. C. Fernandez-Lopez, A. Hidalgo-Miranda, J. Melendez-Zajgla, W. Winckler, K. Ardlie, S. B. Gabriel, M. Meyerson, E. S. Lander, G. Getz, T. R. Golub, L. A. Garraway, and J. R. Grandis. 2011. "The mutational landscape of head and neck squamous cell carcinoma." *Science* 333 (6046):1157–60. doi: 10.1126/science.1208130.

Suganuma, M., H. Fujiki, H. Suguri, S. Yoshizawa, M. Hirota, M. Nakayasu, M. Ojika, K. Wakamatsu, K. Yamada, and T. Sugimura. 1988. "Okadaic acid: an additional non-phorbol-12-tetradecanoate-13-acetate-type tumor promoter." *Proc Natl Acad Sci U S A* 85 (6):1768–71. doi: 10.1073/pnas.85.6.1768.

Sveen, A., S. Kilpinen, A. Ruusulehto, R. A. Lothe, and R. I. Skotheim. 2016. "Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes." *Oncogene* 35 (19):2413–27. doi: 10.1038/onc.2015.318.

Taskesen, E., M. Havermans, K. van Lom, M. A. Sanders, Y. van Norden, E. Bindels, R. Hoogenboezem, M. J. Reinders, M. E. Figueroa, P. J. Valk, B. Löwenberg, A. Melnick, and R. Delwel. 2014. "Two splice-factor mutant leukemia subgroups uncovered at the boundaries of MDS and AML using combined gene expression and DNA-methylation profiling." *Blood* 123 (21):3327–35. doi: 10.1182/blood-2013-07-512855.

ten Klooster, J. P., I. Leeuwen, N. Scheres, E. C. Anthony, and P. L. Hordijk. 2007. "Rac1-induced cell migration requires membrane recruitment of the nuclear oncogene SET." *EMBO J* 26 (2):336–45. doi: 10.1038/sj.emboj.7601518.

Thapa, C. J., T. Haataja, U. Pentikäinen, and P. Permi. 2020. "^{1}H, ^{13}C and ^{15}N NMR chemical shift assignments of cAMP-regulated phosphoprotein-19 and -16 (ARPP-19 and ARPP-16)." *Biomol NMR Assign* 14 (2):227–231. doi: 10.1007/s12104-020-09951-w.

Tirosh, I., A. S. Venteicher, C. Hebert, L. E. Escalante, A. P. Patel, K. Yizhak, J. M. Fisher, C. Rodman, C. Mount, M. G. Filbin, C. Neftel, N. Desai, J. Nyman, B. Izar, C. C. Luo, J. M. Francis, A. A. Patel, M. L. Onozato, N. Riggi, K. J. Livak, D. Gennert, R. Satija, B. V. Nahed, W. T. Curry, R. L. Martuza, R. Mylvaganam, A. J. Iafrate, M. P. Frosch, T. R. Golub, M. N. Rivera, G. Getz, O. Rozenblatt-Rosen, D. P. Cahill, M. Monje, B. E. Bernstein, D. N. Louis, A. Regev, and M. L. Suvà. 2016. "Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma." *Nature* 539 (7628):309–313. doi: 10.1038/nature20123.

Tsai, L. L., C. C. Yu, Y. C. Chang, C. H. Yu, and M. Y. Chou. 2011. "Markedly increased Oct4 and Nanog expression correlates with cisplatin resistance in oral squamous cell carcinoma." *J Oral Pathol Med* 40 (8):621–8. doi: 10.1111/j.1600-0714.2011.01015.x.

Tyner, J. W., C. E. Tognon, D. Bottomly, B. Wilmot, S. E. Kurtz, S. L. Savage, N. Long, A. R. Schultz, E. Traer, M. Abel, A. Agarwal, A. Blucher, U. Borate, J. Bryant, R. Burke, A. Carlos, R. Carpenter, J. Carroll, B. H. Chang, C. Coblentz, A. d'Almeida, R. Cook, A. Danilov, K. T. Dao, M. Degnin, D. Devine, J. Dibb, D. K. Edwards, C. A. Eide, I. English, J. Glover, R. Henson, H. Ho, A. Jemal, K. Johnson, R. Johnson, B. Junio, A. Kaempf, J. Leonard, C. Lin, S. Q. Liu, P. Lo, M. M. Loriaux, S. Luty, T. Macey, J. MacManiman, J. Martinez, M. Mori, D. Nelson, C. Nichols, J. Peters, J. Ramsdill, A. Rofelty, R. Schuff, R. Searles, E. Segerdell, R. L. Smith, S. E. Spurgeon, T. Sweeney, A. Thapa, C. Visser, J. Wagner, K. Watanabe-Smith, K. Werth, J. Wolf, L. White, A. Yates, H. Zhang, C. R. Cogle, R. H. Collins, D. C. Connolly, M. W. Deininger, L. Drusbosky, C. S.

Hourigan, C. T. Jordan, P. Kropf, T. L. Lin, M. E. Martinez, B. C. Medeiros, R. R. Pallapati, D. A. Pollyea, R. T. Swords, J. M. Watts, S. J. Weir, D. L. Wiest, R. M. Winters, S. K. McWeeney, and B. J. Druker. 2018. "Functional genomic landscape of acute myeloid leukaemia." *Nature* 562 (7728):526–531. doi: 10.1038/s41586-018-0623-z.

Van Vlierberghe, P., M. van Grotel, J. Tchinda, C. Lee, H. B. Beverloo, P. J. van der Spek, A. Stubbs, J. Cools, K. Nagata, M. Fornerod, J. Buijs-Gladdines, M. Horstmann, E. R. van Wering, J. Soulier, R. Pieters, and J. P. Meijerink. 2008. "The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia." *Blood* 111 (9):4668–80. doi: 10.1182/blood-2007-09-111872.

Ventelä, S., C. Côme, J. A. Mäkelä, R. M. Hobbs, L. Mannermaa, M. Kallajoki, E. K. Chan, P. P. Pandolfi, J. Toppari, and J. Westermarck. 2012. "CIP2A promotes proliferation of spermatogonial progenitor cells and spermatogenesis in mice." *PLoS One* 7 (3):e33209. doi: 10.1371/journal.pone.0033209.

Ventelä, S., J. A. Mäkelä, J. Kulmala, J. Westermarck, and J. Toppari. 2012. "Identification and regulation of a stage-specific stem cell niche enriched by Nanog-positive spermatogonial stem cells in the mouse testis." *Stem Cells* 30 (5):1008–20. doi: 10.1002/stem.1077.

Vera, J., L. Lartigue, S. Vigneron, G. Gadea, V. Gire, M. Del Rio, I. Soubeyran, F. Chibon, T. Lorca, and A. Castro. 2015. "Greatwall promotes cell transformation by hyperactivating AKT in human malignancies." *Elife* 4. doi: 10.7554/eLife.10115.

Vetrie, D., G. V. Helgason, and M. Copland. 2020. "The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML." *Nat Rev Cancer* 20 (3):158–173. doi: 10.1038/s41568-019-0230-9.

Visvader, J. E., and G. J. Lindeman. 2008. "Cancer stem cells in solid tumours: accumulating evidence and unresolved questions." *Nat Rev Cancer* 8 (10):755–68. doi: 10.1038/nrc2499.

Vitali, C., C. Bassani, C. Chiodoni, E. Fellini, C. Guarnotta, S. Miotti, S. Sangaletti, F. Fuligni, L. De Cecco, P. P. Piccaluga, M. P. Colombo, and C. Tripodo. 2015. "SOCS2 Controls Proliferation and Stemness of Hematopoietic Cells under Stress Conditions and Its Deregulation Marks Unfavorable Acute Leukemias." *Cancer Res* 75 (11):2387–99. doi: 10.1158/0008-5472.CAN-14-3625.

Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler. 2013. "Cancer genome landscapes." *Science* 339 (6127):1546–58. doi: 10.1126/science.1235122.

von Lindern, M., S. van Baal, J. Wiegant, A. Raap, A. Hagemeijer, and G. Grosveld. 1992. "Can, a putative oncogene associated with myeloid leukemogenesis, may be activated by fusion of its 3' half to different genes: characterization of the set gene." *Mol Cell Biol* 12 (8):3346–55. doi: 10.1128/mcb.12.8.3346.

Voso, M. T., T. Ottone, S. Lavorgna, A. Venditti, L. Maurillo, F. Lo-Coco, and F. Buccisano. 2019. "MRD in AML: The Role of New Techniques." *Front Oncol* 9:655. doi: 10.3389/fonc.2019.00655.

Walter, R. B., M. Othus, A. K. Burnett, B. Löwenberg, H. M. Kantarjian, G. J. Ossenkoppele, R. K. Hills, F. Ravandi, T. Pabst, A. Evans, S. R. Pierce, M. C. Vekemans, F. R. Appelbaum, and E. H. Estey. 2015. "Resistance prediction in AML: analysis of 4601 patients from MRC/NCRI, HOVON/SAKK, SWOG and MD Anderson Cancer Center." *Leukemia* 29 (2):312–20. doi: 10.1038/leu.2014.242.

Walter, R. B., M. Othus, A. K. Burnett, B. Löwenberg, H. M. Kantarjian, G. J. Ossenkoppele, R. K. Hills, K. G. van Montfort, F. Ravandi, A. Evans, S. R. Pierce, F. R. Appelbaum, and E. H. Estey. 2013. "Significance of FAB subclassification of "acute myeloid leukemia, NOS" in the 2008 WHO classification: analysis of 5848 newly diagnosed patients." *Blood* 121 (13):2424–31. doi: 10.1182/blood-2012-10-462440.

Wandzioch, E., M. Pusey, A. Werda, S. Bail, A. Bhaskar, M. Nestor, J. J. Yang, and L. M. Rice. 2014. "PME-1 modulates protein phosphatase 2A activity to promote the malignant phenotype of endometrial cancer cells." *Cancer Res* 74 (16):4295–305. doi: 10.1158/0008-5472.CAN-13-3130.

Wang, J., J. Okkeri, K. Pavic, Z. Wang, O. Kauko, T. Halonen, G. Sarek, P. M. Ojala, Z. Rao, W. Xu, and J. Westermarck. 2017. "Oncoprotein CIP2A is stabilized via interaction with tumor suppressor PP2A/B56." *EMBO Rep* 18 (3):437–450. doi: 10.15252/embr.201642788.

Wang, Y., J. Liu, B. O. Huang, Y. M. Xu, J. Li, L. F. Huang, J. Lin, J. Zhang, Q. H. Min, W. M. Yang, and X. Z. Wang. 2015. "Mechanism of alternative splicing and its regulation." *Biomed Rep* 3 (2):152–158. doi: 10.3892/br.2014.407.

Wang, Z., and C. B. Burge. 2008. "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." *RNA* 14 (5):802–13. doi: 10.1261/rna.876308.

Wei, W. J., S. R. Mu, M. Heiner, X. Fu, L. J. Cao, X. F. Gong, A. Bindereif, and J. Hui. 2012. "YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts." *Nucleic Acids Res* 40 (17):8622–36. doi: 10.1093/nar/gks579.

Weichselbaum, R. R., H. Liang, L. Deng, and Y. X. Fu. 2017. "Radiotherapy and immunotherapy: a beneficial liaison?" *Nat Rev Clin Oncol* 14 (6):365–379. doi: 10.1038/nrclinonc.2016.211.

Westermarck, J., and W. C. Hahn. 2008. "Multiple pathways regulated by the tumor suppressor PP2A in transformation." *Trends Mol Med* 14 (4):152–60. doi: 10.1016/j.molmed.2008.02.001.

Wu, C. G., A. Zheng, L. Jiang, M. Rowse, V. Stanevich, H. Chen, Y. Li, K. A. Satyshur, B. Johnson, T. J. Gu, Z. Liu, and Y. Xing. 2017. "Methylation-regulated decommissioning of multimeric PP2A complexes." *Nat Commun* 8 (1):2272. doi: 10.1038/s41467-017-02405-3.

Wunderlich, M., F. S. Chou, K. A. Link, B. Mizukawa, R. L. Perry, M. Carroll, and J. C. Mulloy. 2010. "AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3." *Leukemia* 24 (10):1785–8. doi: 10.1038/leu.2010.158.

Xiao, W., S. Adhikari, U. Dahal, Y. S. Chen, Y. J. Hao, B. F. Sun, H. Y. Sun, A. Li, X. L. Ping, W. Y. Lai, X. Wang, H. L. Ma, C. M. Huang, Y. Yang, N. Huang, G. B. Jiang, H. L. Wang, Q. Zhou, X. J. Wang, Y. L. Zhao, and Y. G. Yang. 2016. "Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing." *Mol Cell* 61 (4):507–519. doi: 10.1016/j.molcel.2016.01.012.

Xing, Y., Z. Li, Y. Chen, J. B. Stock, P. D. Jeffrey, and Y. Shi. 2008. "Structural mechanism of demethylation and inactivation of protein phosphatase 2A." *Cell* 133 (1):154–63. doi: 10.1016/j.cell.2008.02.041.

Xu, Y., Y. Xing, Y. Chen, Y. Chao, Z. Lin, E. Fan, J. W. Yu, S. Strack, P. D. Jeffrey, and Y. Shi. 2006. "Structure of the protein phosphatase 2A holoenzyme." *Cell* 127 (6):1239–51. doi: 10.1016/j.cell.2006.11.033.

Yamamoto, C., H. Nakashima, T. Ikeda, S. I. Kawaguchi, Y. Toda, S. Ito, K. Mashima, T. Nagayama, K. Umino, D. Minakata, H. Nakano, K. Morita, R. Yamasaki, M. Sugimoto, Y. Ishihara, M. Ashizawa, K. Hatano, K. Sato, I. Oh, S. I. Fujiwara, M. Ueda, K. Ohmine, K. Muroi, and Y. Kanda. 2019. "Analysis of the cost-effectiveness of treatment strategies for CML with incorporation of treatment discontinuation." *Blood Adv* 3 (21):3266–3277. doi: 10.1182/bloodadvances.2019000745.

Yang, Y., Q. Huang, Y. Lu, X. Li, and S. Huang. 2012. "Reactivating PP2A by FTY720 as a novel therapy for AML with C-KIT tyrosine kinase domain mutation." *J Cell Biochem* 113 (4):1314–22. doi: 10.1002/jcb.24003.

Yang, Y. T., Y. C. Chiu, C. J. Kao, H. A. Hou, C. C. Lin, C. H. Tsai, M. H. Tseng, W. C. Chou, and H. F. Tien. 2018. "The prognostic significance of global aberrant alternative splicing in patients with myelodysplastic syndrome." *Blood Cancer J* 8 (8):78. doi: 10.1038/s41408-018-0115-2.

Ye, H., Q. Jin, X. Wang, and Y. Li. 2020. "MicroRNA-802 Inhibits Cell Proliferation and Induces Apoptosis in Human Laryngeal Cancer by Targeting cAMP-Regulated Phosphoprotein 19." *Cancer Manag Res* 12:419–430. doi: 10.2147/CMAR.S228429.

Yu, J., A. Boyapati, and K. Rundell. 2001. "Critical role for SV40 small-t antigen in human cell transformation." *Virology* 290 (2):192–8. doi: 10.1006/viro.2001.1204.

Zack, T. I., S. E. Schumacher, S. L. Carter, A. D. Cherniack, G. Saksena, B. Tabak, M. S. Lawrence, C. Z. Zhsng, J. Wala, C. H. Mermel, C. Sougnez, S. B. Gabriel, B. Hernandez, H. Shen, P. W. Laird, G. Getz, M. Meyerson, and R. Beroukhim. 2013. "Pan-cancer patterns of somatic copy number alteration." *Nat Genet* 45 (10):1134–40. doi: 10.1038/ng.2760.

Zagozdzon, R., and J. Golab. 2015. "Cancer stem cells in haematological malignancies." *Contemp Oncol (Pozn)* 19 (1A):A1–6. doi: 10.5114/wo.2014.47127.

UNIVERSITY OF TURKU

Painosalama Oy, Turku, Finland 2021

ISBN 978-951-29-8392-6 (PRINT)
ISBN 978-951-29-8393-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)