**Discovery of NOvel CIP2A VAriant (NOCIVA) and its clinical relevance in myeloid leukemias**

Eleonora Mäkelä1,2,3, Karolina Pavic1,Taru Varila1, Urpu Salmenniemi4, Eliisa Löyttyniemi5, Srikar Nagelli1,6, Tea Ammunét1**,** Veli-Matti Kähäri7, Richard E Clark8,Laura L. Elo1,3, Venkata Kumari Bachanaboyina9,Claire M Lucas8,9, Maija Itälä-Remes10, and Jukka Westermarck1,3,#

### 1Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

### 2Turku Doctoral Programme of Molecular Medicine, University of Turku, Turku, Finland

3Institute of Biomedicine, University of Turku, Turku, Finland

### 4Department of Hematology, Comprehensive Cancer Center, Helsinki University Hospital, Helsinki, Finland

### 5Department of Biostatistics, University of Turku, Turku, Finland

6Drug Research Doctoral Programme, University of Turku, Turku, Finland

### 7Department of Dermatology, University of Turku and Turku University Hospital, Turku, Finland

### 8Department of Molecular, Clinical and Cancer Medicine, University of Liverpool, Liverpool, United Kingdom

### 9Chester Medical School, University of Chester, Chester, United Kingdom

### 10Department of Hematology, Turku University Hospital, Turku, Finland

Running title: Discovery of NOvel CIP2A VAriant (NOCIVA)

Word count: 4920

#Corresponding author: Jukka Westermarck, Turku Bioscience Centre, Tykistökatu 6A, 20540 Turku, tel. +358-29-450 2880, jukwes@bioscience.fi

**Financial support:**

This work was supported by funding from The Sigrid Juselius Foundation (JW), Turku Doctoral Program of Molecular Medicine (EM), University of Turku faculty of medicine (JW, EM), Turku University Hospital ERVA (13283, 13336)(JW, MI-R), The Päivikki and Sakari Sohlbergin Foundation (EM), The Cancer Foundation Väre (EM), The Finnish Cultural Foundation (EM), The Finnish Concordia Fund (EM) and Business Finland TUTL (2445/31/2017)(JW).

**Disclosure of Conflicts of Interest:**

J.W. and E.M have following patents pending related to use of NOCIVA:

“A NOVEL CIP2A VARIANT AND USES THEREOF” (WO/2019/097122) covering use of binding bodies such as probes, amplification primers, and antibodies specific for NOCIVA detection. Also provided are various methods for detecting and prognosing cancer on the basis of said splice variant, and a kit for use in said methods. “METHOD FOR PREDICTING RESPONSE TO TREATMENT WITH TYROSINE KINASE INHIBITORS AND RELATED METHODS” (WO/2020/212650) covering diagnostic evaluation of NOCIVA for predicting response to treatment with tyrosine kinase inhibitors in subjects with haematological cancers, such as CML or AML. In the past 3 years, R.E.C and C.L. declare relevant research support from Bristol Myers Squibb, and R.E.C. declares non-relevant research support and honoraria from Pfizer and Novartis and non-relevant honoraria from Jazz pharmaceuticals and Abbvie.

**Statement of Translational Relevance:** A significant fraction of chronic myeloid leukemia (CML) patients relapse from first-generation (1G) tyrosine kinase inhibitor (TKI) therapy, and this could be prevented by the *de novo* use of 2G TKIs. However, there are no patient stratification markers to guide clinicians in the selection of frontline TKI therapy for optimal outcomes in CML patients. Here, identical results from two independent CML cohorts indicate that the determination of *NOCIVA* mRNA levels by clinically approved qPCR platforms from *de novo* chronic phase CML patients could provide significant support for clinicians in recognizing patients in need of first-line 2G TKI therapy. In acute myeloid leukemia (AML), high *NOCIVA/CIP2A* ratio at diagnosis could indicate a need for first-line therapy intensification.

**Abstract**

Purpose: Cancerous inhibitor of PP2A (CIP2A) is an oncoprotein that inhibits the tumor suppressor PP2A-B56. However, *CIP2A* mRNA variants remain uncharacterized. Here, we report the discovery of a *CIP2A* splicing variant, *NOCIVA* (NOvel CIp2a VAriant).

Experimental Design: Characterization of CIP2A variants was performed by both 3’ and 5’ rapid amplification of cDNA ends from cancer cells. The function of NOCIVA was assessed by structural and molecular biology approaches. Its clinical relevance was studied in an acute myeloid leukemia (AML) patient cohort and two independent chronic myeloid leukemia (CML) cohorts.   
Results: *NOCIVA* contains *CIP2A* exons 1-13 fused to 349 nucleotides from *CIP2A* intron 13. Intriguingly, the first 39 nucleotides of the *NOCIVA*-specific sequence are in the coding frame with exon 13 of *CIP2A* and code for a 13 amino acid peptide tail nonhomologous to any known human protein sequence. Therefore, NOCIVA translates to a unique human protein. NOCIVA retains the capacity to bind to B56, but whereas CIP2A is predominantly a cytoplasmic protein, NOCIVA translocates to the nucleus. Indicative of prevalent alternative splicing from *CIP2A* to *NOCIVA* in myeloid malignancies, AML and CML patient samples overexpress *NOCIVA* but not *CIP2A* mRNA. In AML, a high *NOCIVA/CIP2A* mRNA expression ratio is a marker for adverse overall survival. In CML, high *NOCIVA* expression is associated with inferior event-free survival among imatinib-treated patients, but not among patients treated with dasatinib or nilotinib.

Conclusions: We discovered novel variant of the oncoprotein CIP2A and its clinical relevance in predicting tyrosine kinase inhibitor therapy resistance in myeloid leukemias.

Keywords: KIAA1524, TKI, head and neck squamous cell carcinoma, cutaneous squamous cell carcinoma, cancer, glioblastoma, breast cancer, testis, pancreas, therapy stratification, diagnostic test, quantitative real-time PCR

**Introduction**

Cancerous inhibitor of protein phosphatase 2A (CIP2A) functions as an oncoprotein by directly binding to the tumor suppressor PP2A-B56α (1,2). CIP2A is overexpressed in a vast variety of human cancers, and high CIP2A expression has been shown to correlate with poor patient survival in a broad spectrum of human malignancies (3-6). Furthermore, CIP2A is required for malignant cellular growth *in vitro* and for tumor formation *in vivo* in a number of cancers*,* and its overexpression broadly promotes cancer cell drug resistance (6-11). Prior to the advancement of this potential cancer therapy target in drug development, there should be a comprehensive understanding of CIP2A protein and/or mRNA variants. In regard to the *CIP2A (KIAA1524)* gene, there are no genetic homologs in the human genome, and virtually no information exists about the variant forms of CIP2A at either the mRNA or protein level.

Current evidence suggests a pivotal role of alternative splicing (AS) abnormalities in leukemia pathogenesis (12,13). Particularly in acute myeloid leukemia (AML), a prominent component of the disease is recurrent mutations in spliceosome machinery and genome-wide aberrant splicing events (14-17). Despite therapeutic progress, the outlook for AML remains unsatisfactory (18), and up to 50% of AML patients will experience relapse (18,19). In contrast, chronic myeloid leukemia (CML) treatment was revolutionized by the use of targeted tyrosine kinase inhibitors (TKIs)(20,21). However, despite their remarkable clinical efficiency in most CML patients, a significant portion of patients develop resistance to first-generation (1G) TKI imatinib. The identification of such a patient population at the diagnostic phase is important so that these patients could be treated with first-line 2G TKI therapy instead. However, there are no molecular markers that predict treatment failure at the time of diagnosis or for selection of frontline TKI therapy for optimal outcomes (22). Interestingly, both AML and CML are among very few human malignancies in which *CIP2A* mRNA is not overexpressed; although presumably due to posttranscriptional stabilization, CIP2A is overexpressed at the protein level, and this correlates with more aggressive disease (4,23).

Here, we identified a novel CIP2A variant, NOCIVA (NOvel CIP2A VAriant), that is produced via AS. *NOCIVA* translates to a unique human protein that can heterodimerize with CIP2A, and bind to the PP2A B56 subunit. In AML and CML, high *NOCIVA* expression is found as a marker of poor clinical outcome. Of particular clinical relevance in CML, high *NOCIVA* expression is associated with resistance to the 1G TKI imatinib, but this effect is not seen in patients treated with second-generation TKIs such as dasatinib or nilotinib.

**Materials and Methods**

**3’ RACE and 5’ RACE**

For both 3’ and 5’ rapid amplification of cDNA ends, Invitrogen’s (Carlsbad, CA, USA) 3’RACE (catalog no. 183743-019) and 5’RACE (catalog no. 18374-058) kits were used according to the manufacturer’s protocols. Primer sequences are listed in Supplemental Table 1.

**Quantitative real-time PCR (RQ-PCR)**

The primer and probe sequences used in this study for RQ-PCR analysis are listed in Supplemental Table S1. *NOCIVA* RQ-PCR #1, #2 and #3 assays were designed to amplify the *NOCIVA* specific mRNA sequence (Fig. S1A). The *CIP2A* RQ-PCR assays were designed to amplify the exon13-14 (CIP2A e13 assay) or exon20-21 (CIP2A e20 assay) branch site. The primer concentration in each reaction was 300 nM and probe concentration 200 nM.

The standard curve analysis for amplification efficiency and the melting curve analysis for NOCIVA#1 and NOCIVA#2 RQ-PCRs are shown in Figure S1B-J. The amplification efficiency of all used assays, including control genes beta-actin and GAPDH, was 90-100%. Agarose gel electrophoresis of RQ-PCR products also revealed single band for all main assays used in this study (Fig. S1K). Minor groove binding (MGB) based NOCIVA#3 RQ-PCR assay was used for the analysis of the CML study cohort2 (Fig. S1A).

Amplification of target cDNAs was performed using KAPA PROBE FAST RQ-PCR Kit (Kapa Biosystems, Wilmington, MA, USA) and 7900 HT Fast Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) 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 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 the Thermo Fisher Cloud Real-time qPCR Relative Quantification application. To estimate the degree of overexpression in the AML and CML cohort1, 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 normalised to GAPDH as an endogenous control. Results were derived from the average of at least two independent experiments and two technical replicates.

**Patient cohorts**

The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki.

AML study cohort

Basic molecular and clinical characteristics for the 80 patients with AML are listed in Table S2. Detailed information for AML study cohort can be found in (24). All 80 patients received regimens comprising anthracycline and high-dose cytarabine as induction therapy. Their median age was 50 years (Q1 = 38.8, Q3 = 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 European LeukemiaNet (ELN) 2010 genetic risk group classification (25) was used for risk stratification (Table S3). All patients gave informed, written consent which was signed, and the local Ethical Review Board of Turku University Hospital (TYKS) approved the study protocol.

CML study cohort1

This cohort comprised of 35 newly diagnosed chronic phase CML patients from the University of Liverpool CML biobank. One patient lacked follow up data. Twenty patients received imatinib as a first line therapy and 14 received a second generation TKI, either dasatinib or nilotinib. Their median age was 53.5 years (Q1 = 42.3, Q3 = 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 CML cohort1 patients gave informed, written consent which was signed.

CML study cohort2

This cohort consisted of 159 newly diagnosed CML patients from the UK-wide SPIRIT2 clinical trial (26). The samples were the first 141 biobanked samples 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 (Q1 = 43, Q3 = 63) and median follow-up time was 60 months (range 1–60 months).

The vast majority of SPIRIT2 entrants gave informed consent to donate samples to the SPIRIT2 biobank housed at Molecular Pathology Laboratory at the Imperial College Healthcare NHS Trust (ICHNT), in addition to the informed, written consent which was signed required to enter the trial. The present project was approved by the National Cancer Research Institute CML subgroup, who have ownership of this biobank, and ethical approval was given by the Liverpool East Committee of the UK National Research Ethics Committee.

**Statistical analysis**

Statistical analysis was performed using SAS software (version 9.3, SAS Institute Inc., Cary, NC, USA) or GraphPad Prism (version 8.3., GraphPad Software, San Diego, CA, USA). Normal distribution of the data was tested and if needed transformations were performed. All statistical tests were two-sided and declared significant at a p-value of less than 0.05.

Continuous variables were summarized by descriptive statistics (median, interquartile range and range) while frequencies and percentages were calculated for categorical data. Kruskal-Wallis test, Mann-Whitney u-test, one sample 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.

Univariable survival analysis was based on the Kaplan–Meier method where stratum-specific outcomes were compared using log-rank statistics or on Cox proportional hazards regression model. To adjust for the explanatory variables (diagnosis age, expression levels of NOCIVA, CIP2A, SET, EVI1, WT1, ARPP19, TIPRL and PME1), a Cox proportional hazards regression model was used for multivariable analysis. In multivariable analysis, covariates were entered in a stepwise backward manner.

Overall survival (OS) was defined for all patients measured from the date of diagnosis to the date of death from any cause. Event free survival (EFS) was defined as the time 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, 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 diagnosis of chronic phase to the date of accelerated phase or blast crisis.

**RNA sequencing of AML patient samples**

Eight AML patient bone marrow samples were used for deep RNA sequencing study. Samples with RNA quality numbers ≥9 (Fragment Analyzer (Advanced Analytical Technologies, Heidelberg, Germany)) were selected for RNA library preparation with TruSeq RNA library preparation kit v2 (Illumina, San Diego, CA). Paired-end sequencing with a read length of 100 or 150 bp was performed on an Illumina HiSeq 2500 sequencer, yielding ~114 million read pairs per sample. Image analysis, base calling, and quality check was performed with Illumina data analysis pipelines. The sequencing reads were further trimmed of Illumina adapters using Trimmomatic (version 0.39, (27)). The trimmed reads were then aligned to human reference genome (hg38) allowing for novel junctions using STAR-2.6.1b (28) in two pass mode. The aligned reads were visualized against the known CIP2A gene models and the predicted novel CIP2A isoform (NOCIVA) using the Integrative Genomics Viewer (IGV; (29)). The IGV sashimi plots were drawn to verify the support for the novel junction site between CIP2A exon 13 and 14.

**Antibodies**

Two NOCIVA specific antibodies were generated by immunizing rabbits against NOCIVA specific peptide NNKNTQEAFQVTS by BioGenes GmbH (Berlin, Germany). Antibodies for IF imaging were anti-CIP2A (sc-80659, 1:5000, Santa Cruz Biotechnology, Dallas, TX, USA) and Alexa Fluor-conjugated anti-mouse or rabbit secondary antibodies (488, 555, 1:300, Thermo Fisher Scientific). Antibodies used in NOCIVA binding assays included anti-B56α (sc-136045, 1:5000, Santa Cruz Biotechnology), anti-V5 (E10/V4RR, 1:5000, Thermo Fisher Scientific) and anti-GST (CAB4169, 1:10 000, Thermo Fisher Scientific).

**Recombinant protein binding assays**

CIP2A, NOCIVA and B56 recombinant protein expression, purification and interaction assays were performed as described in (1).

**Results**

**Identification of a Novel CIP2A Variant (NOCIVA) mRNA Isoform**

To identify potential mRNA variants of *CIP2A* (gene alias *KIAA1524*), rapid amplification of cDNA ends PCR assays (3’RACE and 5’RACE) were employed in human cell line mRNA samples (PNT2, MDA-MB-231, HeLa). As a result, a novel CIP2A mRNA splice variant (named here as *NOCIVA*) with alternative exon inclusion was identified (Fig. 1A, Fig. S2A). *NOCIVA* comprised of exons 1-13 of *CIP2A* fused C-terminally to a part of the intron between exons 13 and 14 (Fig. 1A). This 349-nucleotide intronic region (Fig. 1A, Fig. S2B) is normally located within intron 13 of the *CIP2A* gene, more precisely ranging from 108561721 to 108562069 in *Homo sapiens* chromosome 3 (GRCh38.p13 reference, annotation release 109.20200228). As a clear evidence that NOCIVA constitutes a functional mRNA transcript, *NOCIVA* mRNA contains a stop codon followed by a 330-nucleotide 3´UTR with a polyadenylation signal (PAS) AATAAA and a poly(A) tail(Fig. 1B, Fig. S2B).

As an indication that *NOCIVA* mRNA is created by AS, the spliced intron region was found to be flanked by GT and AG dinucleotides (Fig. S2B yellow, GU-AG intron) and the junction site between *CIP2A* and *NOCIVA* contains exonic splicing silencer (ESS) matrices, especially Fas ESS and PESS-octamers (Fig. 1C). Further, binding sites for many splice factors, including YB-1, SRp20, Sam68, SLM2, SRp40, and multiple hnRNPs (including hnRNP K), were predicted in the near vicinity of the junction site by SpliceAid 2 (30) and SFmap (version 1.8)(31) (Fig. 1C).

Validation PCR for full-length *NOCIVA* mRNA expression was conducted in the HeLa cell line with forward primers targeting *CIP2A* exon 1 and reverse primers targeting the *NOCIVA*-specific 3’ end of the mRNA (Fig. 2A and Fig. S3A for PCR assay design). Additionally, validation PCR for *NOCIVA* expression was conducted in multiple cancer cell lines with primers specific to the mRNA sequence that codes for the unique C-terminal portion of *NOCIVA* (Fig. 2B, and Fig. S3B for PCR assay design). The correctly sized PCR products were subsequently sequenced to confirm the *NOCIVA* transcript.

To provide methodology-independent validation of the *NOCIVA* transcript, we analyzed the deep RNA sequencing data of 8 AML patient samples. The sequencing reads were aligned to the human reference genome (hg38), allowing for novel junctions using STAR-2.6.1b in two-pass mode (28). The aligned reads were visualized against the known *CIP2A* mRNA sequence and the *NOCIVA* transcript using the Integrative Genomics Viewer (IGV) (32). Two out of eight analyzed samples had detectable expression of reads aligning with the *NOCIVA-*specificsequence (Fig. 2C, Fig. S3C). Both IGV analysis, and Sashimi plots, verified the existence of a novel junction site resulting in a fusion between CIP2A exon 13 and the *NOCIVA-*specificsequence within intron 13 of the *KIAA1524* gene (Fig. 2C and Fig. S3D).

Together, these results identify *NOCIVA* as a novel, alternatively spliced *CIP2A* variant that is expressed in multiple cancer cell lines. The *NOCIVA* transcript has been assigned a GenBank identifier 2402366.

**Characterization of NOCIVA protein**

Interestingly, in *NOCIVA* mRNA, the 5´ end of the *NOCIVA*-specific intronic sequence is fused in the coding frame with the preceding 3´ end of the *CIP2A* mRNA sequence. After 39 nucleotides, corresponding to 13 amino acids (AAs) (red text in Fig. 1B), the C-terminal tail is followed by a classical stop codon TAA. Therefore, the potential NOCIVA protein consists of 545 AAs that are shared with CIP2A, followed by the NOCIVA-specific peptide sequence NNKNTQEAFQVTS (Fig. 1B). The novel 13 AA peptide sequence in NOCIVA did not match any known protein sequence in the human proteome based on a BLAST homology search (33) (Fig. S4A, BLASTP 2.8.1+, Database: nonredundant (nr) protein sequences). We used the recombinant NOCIVA peptide to generate two affinity chromatography-purified NOCIVA-specific antibodies. The specificity of the antibodies was tested by using bacterially produced NOCIVA and CIP2A proteins. Anti-NOCIVA antibodies specifically recognized NOCIVA but did not recognize CIP2A protein fragments (Fig. 2D, Fig. S4B for NOCIVA ab #2 data). Importantly, the NOCIVA signal could be abolished by using a blocking peptide (Fig. 2D). Additionally, the C-terminal CIP2A antibody did not recognize NOCIVA (Fig. S4C).

Interestingly, whereas CIP2A resided predominantly in the cytoplasm of MDA-MB-231 breast cancer cells as expected (3), endogenous NOCIVA positivity was clearly nuclear (Fig. 2E and Fig. S4D). A similar conclusion could be drawn from GFP fusion overexpression studies in both the MDB-MB-231 (Fig. 2F) and HeLa (Fig. S4E) cell lines. Hence, NOCIVA expresses a novel immunogenic peptide sequence, and constitutes a CIP2A variant protein that is abundantly localized to the nucleus.

To address NOCIVA protein functions, recombinant GST-NOCIVA (CIP2A 1-545+13 AA peptide) and GST-CIP2A 1-560 were compared (Fig. S4F for Coomassie staining) in terms of two functions critical for CIP2A-mediated PP2A modulation: protein homodimerization, and direct binding to the B56α subunit of PP2A (1). Consistent with the location of B56α binding regions in the N-terminal part of CIP2A 1-560 (1), which is identical between NOCIVA and CIP2A, both proteins co-immunoprecipitated B56α with equal efficiency *in vitro* (Fig. 3A). NOCIVA could also competently heterodimerize with CIP2A 1-560, albeit with lower affinity than that seen with CIP2A 1-560 homodimers (Fig. 3B). This can be explained by the partial overlap of the CIP2A-NOCIVA fusion site with the AA region mediating CIP2A homodimerization (1)(Fig. 3C, D), and when compared to CIP2A homodimers, in NOCIVA-CIP2A heterodimers, some of the stabilizing interactions are lost (Fig. 3E).

**NOCIVA expression in normal and cancer cells**

To compare expression levels of *NOCIVA* mRNA in biological samples with those of *CIP2A,* we designed and validated two quantitative real-time PCR (RQ-PCR) assays for both *NOCIVA* (NOCIVA#1 and #2 assays) and *CIP2A* (CIP2A e13 and e20 assays) (Fig. S5A). The primer sequences are listed in Table S1, and if not otherwise indicated, NOCIVA#1 and CIP2A e20 were the mainstay assays. Notably, *CIP2A* and *NOCIVA* RQ-PCR assays were optimized to yield similar amplification efficiencies, allowing a direct comparison between their respective expression levels.

*NOCIVA* showed overall low levels of expression across normal human tissues (Fig. 4A), but consistent with its regulation from the same promoter region as *CIP2A,* theexpression profile across different tissues, including high expression in testis, was comparable to that of *CIP2A* (Fig. 4A and S5B). Although the absolute expression of *NOCIVA* was below 7% of *CIP2A* in all tissues, leukocytes, the kidney, and the pancreas had the highest *NOCIVA/CIP2A* mRNA ratio (Fig. 4B). To address the potential overexpression of *NOCIVA* in cancer, we assessed *NOCIVA* mRNA expression between normal epidermal keratinocytes (Ker, NHEK) and patient-derived cutaneous and head and neck squamous cell carcinoma UT-SCC cells (34,35) in which *CIP2A* was overexpressed (Fig. S5C). *NOCIVA* mRNA also showed significantly elevated expression in SCC samples compared to NHEKs (Fig. 4C, p=0.0001 by Student’s t-test).

Considering that the highest *NOCIVA/CIP2A* ratio was found in lymphoid cells (Fig. 4B), we examined *NOCIVA* expression in lymphoid cancer cells. Relatively higher expression of *NOCIVA* than *CIP2A* was observed in most AML (F36P, Eol-1, Kasumi-1, KG-1, MOLM-13) and CML (K562, Ku812, Meg01) cell lines (Fig. 4D). We confirmed the expression of the NOCIVA protein in Kasumi-1 and KG-1 cells by Western blotting. Consistent with the preferential expression of *NOCIVA* mRNA in lymphoid cells, the expression of NOCIVA protein in AML cells exceeded that of the glioblastoma cell line T98G (Fig. 4E). Encouraged by these results, we validated *NOCIVA* gene expression from a panel of 80 clinical AML (BM) and 35 CML (peripheral blood) samples. Consistent with earlier results (4,24), 96% of AML and 94% of CML patients expressed lower levels of *CIP2A* than normal BM controls pooled from 56 males and females (Fig. 4F, G). However, fully supporting active AS from *CIP2A* to *NOCIVA* in myeloid cancers, 77% of AML and 65% CML samples displayed overexpression of *NOCIVA* compared to BM controls (Fig. 4F, G). Based on Pearson pairwise correlation analysis, we found that *NOCIVA* expression levels in AML samples were significantly correlated among the PP2A inhibitor proteins with *PME1* (r=0.43, p=0.0002), *ARPP19* (r=0.37, p=0.0014), and *SET* (r=0.30, p=0.0104), but not with established AML markers Wilms’ tumor 1 (WT1)(36) and ectopic viral integration site-1 (EVI1)(37)(Fig. 4H).

These results show that *NOCIVA* mRNA is robustly overexpressed in patient AML and CML cells compared to normal BM.

**Clinical relevance of *NOCIVA* expression in diagnostic AML samples**

To understand the potential clinical significance of *NOCIVA* AS, we next analyzed the prognostic significance of *NOCIVA* mRNA expression in 80 AML patients treated with intensive chemotherapy (AML study cohort, Table S2 and (24)). After dividing *NOCIVA* expression into high and low expression groups according to the median (2.18, Q1=1.14, Q3=6.65), Kaplan-Meier estimates revealed that high *NOCIVA* mRNA expression was a strong indicator of shorter overall survival (OS) (Fig. 5A, p=0.022 by log-rank test). Interestingly, low *CIP2A e20* (Fig. 5B, p=0.073 by log-rank test) expression was instead a borderline significant predictor of longer OS, indicating that active AS from *CIP2A* to *NOCIVA* is oncogenic in AML.

An additional analysis of the prognostic role of the studied genes for OS was performed by a multivariable Cox proportional hazards model, which included age at diagnosis and diagnostic mRNA expression levels of *CIP2A e13, CIP2A e20, SET, ARPP19, TIPRL, PME1, EVI1, WT1,* and *NOCIVA.* All the covariates were included in the model as independent predictors. After excluding the nonsignificant markers, age at diagnosis (Fig. 5C, p=0.0013, hazard ratio (HR): 1.07), EVI1 (p=0.0004, HR: 1.27) and *NOCIVA* gene expression (p=0.0205, HR: 1.51) were found to be independent prognostic factors for OS. It was notable that the hazard ratio for *NOCIVA* mRNA expression was even higher than that for *EVI1* expression or diagnosis age, both of which are considered strong predictors of AML outcome in current clinical practice (37,38).

We also analyzed the association of the studied markers with clinical characteristics and risk groups. The expression of *NOCIVA* did not show correlations with any of the clinical characteristics: age, sex, leukocyte or BM blast count, secondary leukemia, or presence/absence of a normal karyotype (p>0.05 in all the analyses, Tables S4 and S5). In regard to genetic risk group associations, neither *NOCIVA* nor *CIP2A* expression levels showed an association with the ELN-2010 risk group (Fig. 5D, p>0.05 by Kruskal-Wallis test). On the other hand, and as expected, *EVI1* mRNA expression at diagnosis was significantly different among the three risk groups, and its expression increased in accordance with the risk group (p=0.005 by Kruskal-Wallis test).

As *CIP2A* and *NOCIVA* mRNA expression levels had opposite roles in predicting AML patient OS (Fig. 5A, B), we further tested the predictive role of the ratio of their expression in the same AML cohort. Importantly, among the four evaluated expression ratio pairs, high *NOCIVA/*low *CIP2A* selectively predicted poor OS (Fig. 5E). When pooling the other three ratios together, the high *NOCIVA/*low *CIP2A* ratio was a significant predictor of inferior OS (Fig. 5F, p=0.042 by log-rank test). These data strongly indicate that rather than the absolute expression of these genes, the ratio between spliced and non-spliced *CIP2A* mRNA is the determining factor for AML survival. This finding could help develop a clinically translatable assay including the measurement of *CIP2A* mRNA as an internal control*.*

Together, these data demonstrate a risk group-independent association between a high *NOCIVA/CIP2A* expression ratio, and a poor clinical outcome among AML patients treated with intensive chemotherapy.

**Clinical relevance of *NOCIVA* expression in diagnostic CML samples**

Next, we evaluated the prognostic significance of *NOCIVA* mRNA expression in 34 newly diagnosed chronic phase (CP) CML patients (CML study cohort 1). Twenty patients received imatinib (1G TKI), and 14 received dasatinib or nilotinib (2G TKI) as first-line therapy. As calculation of OS was not reasonable in this cohort due to only one death at 60 months, Kaplan-Meier estimates were used to analyze the event-free survival (EFS). Importantly, after dividing NOCIVA expression into high and low expression groups according to the median (5.5, Q1=0.20, Q3=20.0), analysis revealed that high *NOCIVA* mRNA expression was associated with significantly shorter EFS (Fig. 6A, p=0.024 by log-rank test).

Although TKIs have revolutionized CML therapy, the identification of patients likely to have resistance to frontline 1G TKI imatinib at diagnosis is still a significant unmet clinical challenge (22). Interestingly, EFS was significantly shorter in the high *NOCIVA* patient group treated with imatinib (Fig. 6B, p=0.004 by log-rank test), but this was not seen among patients treated with 2G TKIs (Fig. 6C, p=0.429 by log-rank test). An analysis of the time to complete molecular response (CMR) was used to assess the depth of a patient’s response, with CMR being the deepest form of response. Patients with high *NOCIVA* expression had a significantly inferior time to CMR (Fig. 6D, p=0.039 by log-rank test). Critically, none of the patients with high levels of *NOCIVA* mRNA at diagnosis achieved CMR. Again, among patients treated with 2G TKIs, no association was found between *NOCIVA* expression and CMR, indicating that 2G TKI therapy may overcome the adverse effect of high *NOCIVA* mRNA expression.

These findings were then validated in an independent cohort of 159 patients (CML study cohort 2) from the SPIRIT2 clinical trial (26). In this cohort, 81 patients had received imatinib and 78 had received dasatinib as first-line therapy. Similar to CML study cohort 1, high *NOCIVA* expression at diagnosis was associated with disease progression exclusively among imatinib-treated patients. Imatinib-treated patients who subsequently progressed to blast crisis had higher expression of *NOCIVA* at diagnosis than patients who did not progress (Fig. 6E, p=0.04 by Mann-Whitney U-test). No significant difference was observed for patients treated with dasatinib (Fig. 6E). Interestingly, imatinib-treated patients with the highest quartile *NOCIVA* expression at diagnosis had significantly inferior freedom from progression (FFP) compared to patients with lower *NOCIVA* expression (Fig. 6F, p=0.039 by log-rank test). Consistent with the results from CML study cohort 1, no association between *NOCIVA* expression and FFP was observed among the dasatinib-treated patients (Fig. 6G). Finally, as a further indication of its independent role as a leukemia oncoprotein, we did not detect a significant association between *NOCIVA* expression and any of the *BRC-ABL* transcript variants from the CML cohort 2 (Fig. 6H).

**Discussion**

Regardless of the thorough profiling of CIP2A expression across human cancers (2,5), the mRNA or protein variants of CIP2A remain uncharacterized. Here we describe a novel clinically relevant CIP2A variant NOCIVA. One of the most interesting features of *NOCIVA* is that it codes for a unique immunogenic C-terminal 13 AA peptide tail (Fig. 1B). As no homologous sequences could be identified to the 13 AA NOCIVA tail in the human proteome, NOCIVA can be considered a novel human protein. Strongly indicative of the alternative cellular functions of CIP2A and NOCIVA, the NOCIVA protein was found to be predominantly nuclear, whereas CIP2A mainly resides in the cytoplasm. However, similar to CIP2A, NOCIVA retains the capability to dimerize and bind to B56, indicating that it functions similarly to CIP2A as a PP2A inhibitor protein (1). Further studies on the differential functional roles of NOCIVA and CIP2A are warranted. In particular, the functional importance of the C-terminal truncation of CIP2A in NOCIVA is an important future topic to be addressed, as currently no molecular functions have been assigned to the C-terminal CIP2A tail. Unfortunately, during the project, we failed to yet develop siRNA or CRISPR/Cas9 tools selectively suppressing *NOCIVA* but not *CIP2A*.

AML and CML patient samples displayed clearly higher expression of *NOCIVA* mRNA than *CIP2A*, suggesting that during myeloid leukemogenesis, a splicing switch that creates *NOCIVA* from *CIP2A* is activated (Fig. 6I). This is interesting, as AML and CML are the only cancer types in which CIP2A seems to be underexpressed at the mRNA level compared to corresponding normal tissue (Fig. 6I)(4,23). At the *NOCIVA* junction site, exonic splicing silencer sequences (ESSs) as well as binding sites for hnRNPs and various splice factors were found (Fig. 1C). A recent study reported expression changes in 13 hnRNPs affecting mRNA processing in AML (39), among which hnRNP A1, A2B1, and C were predicted to bind to the *NOCIVA* junction site. Additionally, the expression of hnRNP K (40), SRSF3 (SRp20) (41), and YB-1 (42) has been shown to be altered in AML but also to contribute to leukemia progression. Interestingly, SRSF3 (43) and YB-1 (44) have also been shown to specifically promote exon inclusion during AS. Detailed analysis of the role of these splicing factors in the AS of *CIP2A* to *NOCIVA* will be needed in the future to better understand the regulation of NOCIVA in myeloid cancers.

Our data firmly indicate that high *NOCIVA* mRNA expression is associated with poor clinical outcomes in both AML and CML. The data suggest that if NOCIVA is highly expressed in AML cells at the diagnosis phase, treatment of those patients with standard cytotoxic chemotherapy is not sufficient to kill these cells. Thus, we cautiously propose that NOCIVA contributes to cytotoxic chemotherapy resistance in AML. These novel findings are particularly interesting because *NOCIVA* expression was independent of the current genetic risk classification in AML, suggesting that the evaluation of *NOCIVA* expression at diagnosis could provide clinically relevant predictive value. Furthermore, mechanistic follow-up studies of these findings could reveal novel mechanisms mediating AML chemotherapy resistance. In CML, high *NOCIVA* mRNA expression was associated with inferior EFS and FFP as well as lower rates of CMR in imatinib-treated CML patients. Hence, the data suggest that 2G TKI therapy is needed to overcome the adverse effects of high *NOCIVA* expression. Currently, a significant number of CML patients still receive 1G TKI as front-line therapy due to the higher costs and worse side-effect profiles of 2G and 3G TKIs. We propose that together with other diagnostic biomarkers, such as *BCR-ABL* transcript variants, the detection of *NOCIVA/CIP2A* ratioat the CML diagnostic phase might help in treatment decision making between imatinib and 2G TKIs. The urgent clinical utility of such an assay for CML patient front-line TKI therapy selection was recently highlighted by an expert group (22).

Although previously debated, mRNA expression-based diagnostics have entered clinical CML diagnostics (45). As an example, *BCR-ABL1/ABL1* mRNA ratios on the international scale (IS) were shown to predict the success of treatment-free remission attempts for patients with *de novo* CP CML (46). The technical platforms allowing a standardized clinical assessment of mRNA expression levels for CML samples include droplet digital PCR (ddPCR) (46) and the Cepheid GeneXpert qPCR cartridge system (47). However, currently, there are no patient stratification markers to guide clinicians in the selection of frontline TKI therapy to achieve optimal outcomes in CML patients (22). Based on our nearly identical results from two independent clinical CML cohorts, and independence of *NOCIVA* from other candidate markers, the determination of NOCIVA mRNA levels from patients with *de novo* CP CML by ddPCR assay or the Cepheid qPCR cartridge system could provide significant support for clinicians in recognizing patients in need of first-line 2G TKI therapy. Similarly, in AML, high NOCIVA expression at diagnosis could indicate a need for frontline therapy intensification.

In summary, this work describes the discovery of a novel human gene and protein product with the characteristics of a clinically relevant PP2A inhibitor in myeloid malignancies. Further studies to validate the clinical diagnostic value of *NOCIVA* mRNA in the identification of imatinib-resistant CML patients, and the mechanistic basis for imatinib resistance are clearly warranted.

**Acknowledgements**

Taina Kalevo-Mattila is acknowledged for superior technical assistance. Dr. Veli Kairisto (Turku University Hospital) is acknowledged for his valuable contribution to AML mRNA samples. We gratefully acknowledge the CML subgroup of the United Kingdom National Cancer Research Institute, especially Prof. Jane Apperley, and Sandra Loaiza for access to the SPIRIT2 CML samples and Newcastle University for supplying data from the SPIRIT2 trial. We thank Prof. Maria D. Odero, Drs. Otto Kauko, Juha Okkeri and Mikko Frilander for constructive discussions and advice.

**Authorship Contributions**

E.M. and J.W. conceived the study and experiments; E.M., K.P., T.V., S. N., V.K.B. and C.L. performed the experiments; E.M., E.L. and C.L. analyzed the data; U.S. and M.I-R. collected samples and data from AML patients; R.E.C and C.L collected samples and clinical data from CML patients; V-M. K. provided HNSCC & NHEK cDNA panel; E.M. wrote the manuscript, with input from J.W., K.P., U.S., C.L., R.E.C. and M.I-R. All authors reviewed and approved the final manuscript.

**References**

1. Wang J, Okkeri J, Pavic K, Wang Z, Kauko O, Halonen T*, et al.* Oncoprotein CIP2A is stabilized via interaction with tumor suppressor PP2A/B56. *EMBO reports* 2017;**18**(3):437-50 doi 10.15252/embr.201642788.

2. Soofiyani SR, Hejazi MS, Baradaran B. The role of CIP2A in cancer: A review and update. *Biomed Pharmacother* 2017;**96**:626-33 doi 10.1016/j.biopha.2017.08.146.

3. Junttila MR, Puustinen P, Niemelä M, Ahola R, Arnold H, Böttzauw T*, et al.* CIP2A inhibits PP2A in human malignancies. *Cell* 2007;**130**(1):51-62 doi 10.1016/j.cell.2007.04.044.

4. Lucas CM, Harris RJ, Giannoudis A, Copland M, Slupsky JR, Clark RE. Cancerous inhibitor of PP2A (CIP2A) at diagnosis of chronic myeloid leukemia is a critical determinant of disease progression. *Blood* 2011;**117**(24):6660-8 doi 10.1182/blood-2010-08-304477.

5. Khanna A, Pimanda JE. Clinical significance of Cancerous Inhibitor of Protein Phosphatase 2A (CIP2A) in human cancers. *Int J Cancer* 2015 doi 10.1002/ijc.29431.

6. Côme C, Laine A, Chanrion M, Edgren H, Mattila E, Liu X*, et al.* CIP2A is associated with human breast cancer aggressivity. *Clin Cancer Res* 2009;**15**(16):5092-100 doi 10.1158/1078-0432.CCR-08-3283.

7. Elgendy M, Ciro M, Hosseini A, Weiszmann J, Mazzarella L, Ferrari E*, et al.* Combination of Hypoglycemia and Metformin Impairs Tumor Metabolic Plasticity and Growth by Modulating the PP2A-GSK3beta-MCL-1 Axis. *Cancer Cell* 2019;**35**(5):798-815 e5 doi 10.1016/j.ccell.2019.03.007.

8. Laine A, Sihto H, Come C, Rosenfeldt MT, Zwolinska A, Niemelä M*, et al.* Senescence sensitivity of breast cancer cells is defined by positive feedback loop between CIP2A and E2F1. *Cancer Discov* 2013;**3**(2):182-97 doi 10.1158/2159-8290.CD-12-0292.

9. Khanna A, Böckelman C, Hemmes A, Junttila MR, Wiksten JP, Lundin M*, et al.* MYC-dependent regulation and prognostic role of CIP2A in gastric cancer. *J Natl Cancer Inst* 2009;**101**(11):793-805 doi 10.1093/jnci/djp103.

10. Kauko O, Imanishi SY, Kulesskiy E, Yetukuri L, Laajala TD, Sharma M*, et al.* Phosphoproteome and drug-response effects mediated by the three protein phosphatase 2A inhibitor proteins CIP2A, SET, and PME-1. *J Biol Chem* 2020;**295**(13):4194-211 doi 10.1074/jbc.RA119.011265.

11. Kauko O, O'Connor CM, Kulesskiy E, Sangodkar J, Aakula A, Izadmehr S*, et al.* PP2A inhibition is a druggable MEK inhibitor resistance mechanism in KRAS-mutant lung cancer cells. *Science translational medicine* 2018;**10**(450) doi 10.1126/scitranslmed.aaq1093.

12. Yang YT, Chiu YC, Kao CJ, Hou HA, Lin CC, Tsai CH*, et al.* The prognostic significance of global aberrant alternative splicing in patients with myelodysplastic syndrome. *Blood Cancer J* 2018;**8**(8):78 doi 10.1038/s41408-018-0115-2.

13. Adamia S, Haibe-Kains B, Pilarski PM, Bar-Natan M, Pevzner S, Avet-Loiseau H*, et al.* A genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets. *Clin Cancer Res* 2014;**20**(5):1135-45 doi 10.1158/1078-0432.CCR-13-0956.

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

15. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL*, et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* 2015;**125**(9):1367-76 doi 10.1182/blood-2014-11-610543.

16. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND*, et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016;**374**(23):2209-21 doi 10.1056/NEJMoa1516192.

17. Taskesen E, Havermans M, van Lom K, Sanders MA, van Norden Y, Bindels E*, et al.* Two splice-factor mutant leukemia subgroups uncovered at the boundaries of MDS and AML using combined gene expression and DNA-methylation profiling. *Blood* 2014;**123**(21):3327-35 doi 10.1182/blood-2013-07-512855.

18. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med* 2015;**373**(12):1136-52 doi 10.1056/NEJMra1406184.

19. Petersdorf SH, Kopecky KJ, Slovak M, Willman C, Nevill T, Brandwein J*, et al.* A phase 3 study of gemtuzumab ozogamicin during induction and postconsolidation therapy in younger patients with acute myeloid leukemia. *Blood* 2013;**121**(24):4854-60 doi 10.1182/blood-2013-01-466706.

20. Cross NC, White HE, Colomer D, Ehrencrona H, Foroni L, Gottardi E*, et al.* Laboratory recommendations for scoring deep molecular responses following treatment for chronic myeloid leukemia. *Leukemia* 2015;**29**(5):999-1003 doi 10.1038/leu.2015.29.

21. Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley JF*, et al.* European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. *Blood* 2013;**122**(6):872-84 doi 10.1182/blood-2013-05-501569.

22. Branford S, Kim DDH, Apperley JF, Eide CA, Mustjoki S, Ong ST*, et al.* Laying the foundation for genomically-based risk assessment in chronic myeloid leukemia. *Leukemia* 2019;**33**(8):1835-50 doi 10.1038/s41375-019-0512-y.

23. Barragán E, Chillón MC, Castelló-Cros R, Marcotegui N, Prieto MI, Hoyos M*, et al.* CIP2A high expression is a poor prognostic factor in normal karyotype acute myeloid leukemia. *Haematologica* 2015;**100**(5):e183-5 doi 10.3324/haematol.2014.118117.

24. Mäkelä E, Löyttyniemi E, Salmenniemi U, Kauko O, Varila T, Kairisto V*, et al.* Arpp19 Promotes Myc and Cip2a Expression and Associates with Patient Relapse in Acute Myeloid Leukemia. *Cancers (Basel)* 2019;**11**(11) doi 10.3390/cancers11111774.

25. Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK*, et al.* Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 2010;**115**(3):453-74 doi 10.1182/blood-2009-07-235358.

26. O’Brien S, Cork L, Bandiera V, Bescoby R, Foroni L, Alaily L*, et al.* 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. Volume 132: Blood; 2018. p (Supplement 1): 457.

27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 2014;**30**(15):2114-20 doi 10.1093/bioinformatics/btu170.

28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* 2013;**29**(1):15-21 doi 10.1093/bioinformatics/bts635.

29. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G*, et al.* Integrative genomics viewer. *Nature biotechnology* 2011;**29**(1):24-6 doi 10.1038/nbt.1754.

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

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

32. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;**14**(2):178-92 doi 10.1093/bib/bbs017.

33. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W*, et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**(17):3389-402.

34. Johansson N, Saarialho-Kere U, Airola K, Herva R, Nissinen L, Westermarck J*, et al.* Collagenase-3 (MMP-13) is expressed by hypertrophic chondrocytes, periosteal cells, and osteoblasts during human fetal bone development. *Dev Dyn* 1997;**208**:387-95.

35. Farshchian M, Kivisaari A, Ala-Aho R, Riihila P, Kallajoki M, Grenman R*, et al.* Serpin peptidase inhibitor clade A member 1 (SerpinA1) is a novel biomarker for progression of cutaneous squamous cell carcinoma. *Am J Pathol* 2011;**179**(3):1110-9 doi 10.1016/j.ajpath.2011.05.012.

36. Ujj Z, Buglyó G, Udvardy M, Beyer D, Vargha G, Biró S*, et al.* WT1 Expression in Adult Acute Myeloid Leukemia: Assessing its Presence, Magnitude and Temporal Changes as Prognostic Factors. *Pathol Oncol Res* 2016;**22**(1):217-21 doi 10.1007/s12253-015-0002-0.

37. Hinai AA, Valk PJ. Review: Aberrant EVI1 expression in acute myeloid leukaemia. *Br J Haematol* 2016;**172**(6):870-8 doi 10.1111/bjh.13898.

38. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T*, et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;**129**(4):424-47 doi 10.1182/blood-2016-08-733196.

39. Alanazi B, Munje CR, Rastogi N, Williamson AJK, Taylor S, Hole PS*, et al.* Integrated nuclear proteomics and transcriptomics identifies S100A4 as a therapeutic target in acute myeloid leukemia. *Leukemia* 2020;**34**(2):427-40 doi 10.1038/s41375-019-0596-4.

40. Gallardo M, Lee HJ, Zhang X, Bueso-Ramos C, Pageon LR, McArthur M*, et al.* hnRNP K Is a Haploinsufficient Tumor Suppressor that Regulates Proliferation and Differentiation Programs in Hematologic Malignancies. *Cancer Cell* 2015;**28**(4):486-99 doi 10.1016/j.ccell.2015.09.001.

41. Liu J, Huang B, Xiao Y, Xiong HM, Li J, Feng DQ*, et al.* Aberrant expression of splicing factors in newly diagnosed acute myeloid leukemia. *Onkologie* 2012;**35**(6):335-40 doi 10.1159/000338941.

42. Liu S, Marneth AE, Alexe G, Walker SR, Gandler HI, Ye DQ*, et al.* The kinases IKBKE and TBK1 regulate MYC-dependent survival pathways through YB-1 in AML and are targets for therapy. *Blood Adv* 2018;**2**(23):3428-42 doi 10.1182/bloodadvances.2018016733.

43. Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF*, et al.* Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* 2016;**61**(4):507-19 doi 10.1016/j.molcel.2016.01.012.

44. Wei WJ, Mu SR, Heiner M, Fu X, Cao LJ, Gong XF*, et al.* YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic Acids Res* 2012;**40**(17):8622-36 doi 10.1093/nar/gks579.

45. Radich JP, Deininger M, Abboud CN, Altman JK, Berman E, Bhatia R*, et al.* Chronic Myeloid Leukemia, Version 1.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2018;**16**(9):1108-35 doi 10.6004/jnccn.2018.0071.

46. Nicolini FE, Dulucq S, Boureau L, Cony-Makhoul P, Charbonnier A, Escoffre-Barbe M*, et al.* 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* 2019;**25**(22):6606-13 doi 10.1158/1078-0432.CCR-18-3373.

47. Gerrard G, Foong HE, Mudge K, Alikian M, Apperley JF, Foroni L. Cepheid xpert monitor platform for the confirmation of BCR-ABL1 IS conversion factors for the molecular monitoring of chronic myeloid leukaemia. *Leukemia research* 2016;**49**:47-50 doi 10.1016/j.leukres.2016.08.007.

**Figure legends**

**Figure 1. Characterization of NOvel CIP2A VAriant (NOCIVA) mRNA isoform**

**A)** A schematic presentation of the *NOCIVA* mRNA isoform identified with RACE-PCR. *NOCIVA* mRNA contains an alternative exon from *CIP2A* intron 13, and thus forms a unique and previously unknown coding sequence. Untranslated regions (5’UTR or 3’UTR) are marked with dots, the unique alternative exon in *NOCIVA* with red and *NOCIVA* specific 3’UTR with blue. Full length *CIP2A* corresponds to RefSeq NM\_020890.2 sequence. **B)** *NOCIVA* mRNA’s 3’-end with the differing features from the original *CIP2A* mRNA sequence. The shared nucleotide sequence between *CIP2A* and *NOCIVA* mRNA is underlined. NOCIVA protein comprises 545 N-terminal CIP2A amino acids and 13 unique amino acids on the C-terminus (in red). The stop codon is indicated by an asterisk. *NOCIVA* mRNA contains a 3’UTR (blue, 1675-2010) with a polyadenylation signal (PAS, AATAAA at 1962-1967). **C)** The *NOCIVA* splice junction with splice site predictions from SpliceAid 2.

**Figure 2. NOCIVA is a novel CIP2A splice variant that constitutes a CIP2A variant protein expressed abundantly in the nucleus**

**A)** PCR validation of the *NOCIVA* full-length mRNA sequence expression from the Hela cell line. The forward primer for all lanes was the same targeting the *CIP2A* exon1 and the distinct reverse primers (R1, R2, R3) were all *NOCIVA* specific. Please see Fig. S3A for schematic presentation of the PCR assay design. Arrows indicate the PCR product with expected size. The indicated bands were extracted and the presence of *NOCIVA* specific cDNA (mRNA) was confirmed by DNA sequencing. **B)** PCR validation of *NOCIVA* specific mRNA expression from several cancer cell lines. PCR was conducted with *CIP2A* exon13 targeting forward primer and the *NOCIVA* sequence specific reverse primer. Please see Fig. S3B for schematic presentation of the PCR assay design. NTC = non-template control. **C)** RNA sequencing alignment data of one AML patient sample for the NOCIVA specific mRNA transcipt. The sequencing reads were aligned to human reference genome (HG38). The aligned reads were visualized against the known *CIP2A* mRNA sequence and the *NOCIVA* transcript using the Integrative Genomics Viewer (IGV). **D)** NOCIVA specific antibody detects correct size (appr. 90 kDa) recombinant GST-NOCIVA protein, but not recombinant CIP2A fragments. One microgram of each protein was loaded. The signal was blocked with a NOCIVA specific peptide. Above the blots is a schematic presentation of the different CIP2A and NOCIVA fragments used. Full length CIP2A comprises of 905 aa. **E)** Representative confocal immunofluorescence images of MDA-MB-231 cells stained with NOCIVA and CIP2A antibodies for endogenous proteins. NOCIVA green, CIP2A red, nucleus blue. Scale bar 10 µm. **F)** Representative confocal images of MDA-MB-231 cells transiently overexpressed with NOCIVA-GFP plasmid for 48 hours. Two separate fields are shown. NOCIVA-GFP green, nucleus blue. Scale bar 10 µm.

**Figure 3. Characterization of NOCIVA protein**

**A)** *In vitro* GST‐pulldown assay for interaction between B56α and GST‐CIP2A(1–560), or GST-NOCIVA. Representative image from three experiments is shown. The graph shows relative B56α ‐binding efficiency of GST-NOCIVA as compared to GST‐CIP2A(1–560), quantified as a ratio between B56α and GST‐CIP2A(1–560) in pulldown samples. Each bar is mean ± SD from three independent B56α ‐binding experiments; p = 0.405 by one sample t-test. **B)** *In vitro* hetero-dimerization assay using purified recombinant GST-tagged NOCIVA and CIP2A(1–560) proteins. Representative image from three experiments is shown. The graph shows relative dimerization efficiency of GST-NOCIVA - V5-CIP2A (1-560) heterodimer as compared to GST‐CIP2A (1–560) - V5-CIP2A (1-560) homodimer, quantified as a ratio between CIP2A(1–560)‐V5 and GST‐CIP2A(1–560) from a pulldown sample. Each bar is mean ± SD from three independent experiments; p = 0.017 by one sample t-test. **C)** Crystal structure of CIP2A(1-560) homodimer (PDB: 5UFL). **D)** Modeling of dimer interface area of CIP2A-NOCIVA heterodimer. Differences in NOCIVA (residues 546-560), in contrast to CIP2A, are mapped on CIP2A´s surface and shown in purple-blue.C and D were generated in The PyMOL Molecular Graphics System (Version 2.0 Schrödinger, LCC). **E)** Amino acid residues distinct between CIP2A(1-560) (left panel) and NOCIVA (right panel) are indicated as sticks and colored based on heteroatom. Protein molecule orientation was held in approximately the same for both panels but twisted slightly to show the optimal orientation of the key residues. Image was done using UCSF Chimera (Version 1.14). Differences in the nature of amino acid side chains are represented by the color scheme and also indicated in the alignment, following the same coloring pattern.

**Figure 4. *NOCIVA* mRNA expression in normal and cancerous cells**

**A)** *NOCIVA* mRNA expression in normal tissue panel (Human MTC panel I & II, Clontech) measured with NOCIVA #1 RQ-PCR assay. **B)** *NOCIVA*/*CIP2A* mRNA expression ratio in normal tissues. The percentage value was obtained from a ratio between NOCIVA expression from A) and CIP2A expression from Figure S5B. **C)** *NOCIVA* mRNA expression in patient derived normal human epidermal keratinocytes (Ker, NHEK) and squamous cell carcinoma (SCC) cells. mRNA expression levels in Ker 45B cells were defined as value 1. RQ = relative quantification. p = 0.0001 by Student’s t-test**. D)** *NOCIVA* and *CIP2A* mRNA expression in AML and CML cell lines. mRNA expression levels in Hela cells were defined as value 1. **E)** Representative western blot images of NOCIVA and CIP2A protein expression in indicated AML and glioblastoma (T98G) cell lines. Three independent cell samples were used from each AML cell line. **F)** and **G)** Waterfall plots of analyzed genes from the patient cohorts normalized to the pooled (n=56) normal BM sample. On the y-axis are log10 transformed RQ mRNA expression values derived from two technical replicates in two independent experiments. One bar represents one patient. **H)** Pearson’s pairwise correlations for the mRNA expression of PP2A inhibitors in the AML patient cohort. *NOCIVA* correlates with *PME1* (r=0.43, p=0.0002), *ARPP19* (r=0.37, p=0.0014) and *SET* (r=0.30, p=0.0104), but not with other studied markers. Red represents positive and blue negative correlation. Grey indicates non-significant correlation (p-value > 0.05).

*Beta-actin* & *GAPDH* were used as housekeeping genes in all experiments presented in this figure. Expression values are derived from three technical replicates in two independent experiments. All the figures show mean ± standard error of mean (SEM).

**Figure 5. High *NOCIVA* expression associates with inferior overall survival in AML patients**

**A)** Kaplan–Meier survival curve for overall survival (OS) by *NOCIVA* gene expression in the AML study cohort, stratified according to the median expression. Higher *NOCIVA* expression is associated with shorter OS; p = 0.022 by log-rank test. **B)** Association of *CIP2A* gene expression level in the AML study cohort, stratified according to the median expression, and OS (p = 0.073 by log-rank test). **C)** Multivariable Cox’s proportional hazard model for OS revealed that age at diagnosis (p = 0.0013, HR: 1.07), EVI1 (p = 0.0004, HR: 1.27) and *NOCIVA* (p = 0.0205, HR: 1.51) gene expressions were independent prognostic factors for OS. **D)** Gene expression correlation of the indicated genes with ELN-2010 genetic risk groups in AML patient cohort by Kruskal-Wallis test. Group 1=favourable, 2=intermediate, 3=adverse. **E)** Kaplan–Meier survival curve for OS by *NOCIVA/CIP2A* expression ratio in AML patient cohort. High *NOCIVA/*low *CIP2A* expressing patients associates with inferior OS. **F)** Kaplan–Meier survival curve for OS by *NOCIVA/CIP2A* expression ratio in AML patient cohort. Shown is NOCIVA high/CIP2A low group in relation to other pooled *NOCIVA/CIP2A* ratios (NOCIVA high/CIP2A high, NOCIVA low/CIP2A low, NOCIVA low/CIP2A high). NOCIVA high*/*CIP2A low patients associates with inferior OS; p = 0.042 by log-rank test.

**Figure 6. High *NOCIVA* expression associates with inferior outcome in CML patients treated with imatinib**

**A)** Kaplan–Meier survival curve for event free survival (EFS) by *NOCIVA* mRNA expression in CML study cohort1. The median level of NOCIVA mRNA expression was used to define the high and low groups in each panel. Higher *NOCIVA* expression is associated with shorter EFS; p = 0.024 by log-rank test. **B)** Higher *NOCIVA* expression is associated with shorter EFS in imatinib treated patients in CML cohort1; p = 0.004 by log-rank test. **C)** No significant association was found related to *NOCIVA* gene expression level and EFS in patients treated with 2G TKI in CML cohort1; p = 0.429 by log-rank test. **D)** Lower *NOCIVA* mRNA expression is associated with shorter time to complete molecular response (CMR) in imatinib treated patients in CML cohort1; p = 0.039 by log-rank test. **E)** High *NOCIVA* expression at diagnosis is associated with disease progression for imatinib-treated patients in CML study cohort2; p = 0.04 by Mann-Whitney u-test. Data represents mean ± SEM. **F)** High *NOCIVA* expression is associated with shorter freedom from progression (FFP) in imatinib treated patients in CML cohort2; p = 0.039 by log-rank test. Q4 = highest quartile of *NOCIVA* mRNA expression. **G)** No significant association was found related to *NOCIVA* gene expression level and FFP in patients treated with dasatinib in CML cohort2; p = 0.863 by log-rank test. Q4 = highest quartile of *NOCIVA* mRNA expression. **H)** *BRC-ABL1* transcript variant (E13A2, E14A2, E13A2 and E14A2) expression in association to *NOCIVA* expression in CML study cohort2. No significant association was detected between any *BCR-ABL* transcript variant and *NOCIVA* expression. **I)** Schematic presentation of the distict roles of CIP2A and NOCIVA in CML.Whereas NOCIVA is overexpressed in CML at mRNA level, CIP2A is only overexpressed at protein level. In CP CML, high NOCIVA mRNA level predicts for disease progression after 1G TKI (imatinib) treatment whereas high CIP2A protein level predicts for disease progression after 1G or 2G TKI treatment.