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

Eleonora Mäkelä[1,2,3], Karolina Pavic[1], Taru Varila[1], Urpu Salmenniemi[4], Eliisa Löyttyniemi[5], Srikar Nagelli[1,6], Veli-Matti Kähäri[7], Richard E Clark[8], Venkata Kumari Bachanaboyina[9], Claire Lucas[8,9], Maija Itälä-Remes[10], and Jukka Westermarck[1,3,#]

[1]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland
[2]Turku Doctoral Programme of Molecular Medicine, University of Turku, Turku, Finland
[3]Institute of Biomedicine, University of Turku, Turku, Finland
[4]Department of Hematology, Comprehensive Cancer Center, Helsinki University Hospital, Helsinki, Finland
[5]Department of Biostatistics, University of Turku, Turku, Finland
[6]Drug Research Doctoral Programme, University of Turku, Turku, Finland
[7]Department of Dermatology, University of Turku and Turku University Hospital, Turku, Finland
[8]Department of Molecular, Clinical and Cancer Medicine, University of Liverpool, Liverpool, United Kingdom
[9]Chester Medical School, University of Chester, Chester, United Kingdom
[10]Department of Hematology, Turku University Hospital, Turku, Finland

Word count: 3998 - counting only the Introduction, Methods, Results, and Discussion, Abstract word count: 199, Number of figures and tables: 6, Number of references: 47

Running title: Discovery of NOvel CIP2A VAriant (NOCIVA)

[#]Corresponding author: Jukka Westermarck, tel. +358-29-450 2880, jukwes@utu.fi

Key Points

- Discovery and characterization of a first mRNA variant of one of the most prevalently deregulated human oncoproteins CIP2A
- Unlike CIP2A, NOCIVA mRNA is overexpressed in AML and CML patient samples and associates with poor clinical response in both myeloid cancers

Abstract

Cancerous inhibitor of PP2A (CIP2A) is a prevalent human oncoprotein that inhibits tumor suppressor PP2A-B56a. However, *CIP2A* mRNA and protein variants remain uncharacterized. Here, we report discovery of a *CIP2A* splicing variant *NOCIVA* (NOvel CIp2a VAriant). *NOCIVA* contains *CIP2A* exons 1-13 fused to a continuous stretch of 349 nucleotide from *CIP2A* intron 13. Intriguingly, the first 39 nucleotides of the *NOCIVA* specific sequence are in coding frame with exon 13 of *CIP2A,* and codes for a 13 amino acid peptide tail unhomologous to any known human protein sequence. Therefore, NOCIVA translates to a unique human protein. NOCIVA retains the capacity to bind to B56a, but whereas CIP2A is predominantly a cytoplasmic protein, NOCIVA translocates to nucleus. Indicative of prevalent alternative splicing from *CIP2A* to *NOCIVA* in myeloid malignancies, acute myeloid leukemia (AML) and chronic myeloid leukemia (CML) patient samples overexpress *NOCIVA,* but not *CIP2A* mRNA. In AML, high *NOCIVA* mRNA expression is a marker for adverse overall survival. In CML, high *NOCIVA* expression associates with inferior event free survival among imatinib treated patients, but not among patients treated with dasatinib or nilotinib. Collectively, we describe discovery of a novel variant of oncoprotein CIP2A, and its clinical relevance in myeloid leukemias.

Keywords: CIP2A, KIAA1524, Alternative splicing, splice variant, AML, CML, TKI

Introduction

Cancerous inhibitor of protein phosphatase 2A (CIP2A) functions as an oncoprotein by directly binding to 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 [3-5]. Furthermore, CIP2A is required for malignant cellular growth *in vitro* and for tumor formation *in vivo* in a number of cancers, and its overexpression promotes broadly cancer cell drug resistance [3, 5-10]. However, CIP2A deficiency does not compromise normal mouse development or growth [6, 11, 12]. Consequently, inhibition of CIP2A protein expression or activity could constitute a very efficient cancer therapy strategy with minor side effects. Prior to advancement of this potential cancer therapy target in drug development, and for the design of highly specific therapeutics, 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 apart from *CIP2A* variant database predictions that lack experimental validation, virtually no information exists about variant forms of CIP2A at either mRNA or protein level.

Alternative splicing (AS) is a physiological phenomenon that greatly diversifies the repertoire of the transcriptome. As up to 95% of multi-exonic genes are alternatively spliced [13], AS ensures higher protein diversity for better environmental fit. Perturbation of AS by spliceosome gene mutations, epigenetic modifications, or other causes lead to aberrant AS, and this has

been shown to be a hallmark of cancer development [14]. Pan-cancer studies have revealed that tumors have an average of 20% more AS events than healthy samples [15, 16]. Current evidence suggests a pivotal role of AS abnormalities especially in leukemia pathogenesis, particularly in myelodysplastic syndrome and acute myeloid leukemia (AML) [17, 18].

Myeloid leukemias, including AML and chronic myeloid leukemia (CML), are heterogeneous clonal hematological malignancies that disrupts normal hematopoiesis. Whereas AML is the most common acute leukemia affecting adults, CML accounts for 15–25% of all adult leukemias [19, 20]. Interestingly, both AML and CML are one of the very few human malignancies in which *CIP2A* mRNA is not overexpressed, although presumably due to post-transcriptional stabilization, CIP2A is overexpressed at protein level and this correlates with more aggressive disease [21, 22]. Despite therapeutic progress, the outlook for AML remains unsatisfactory [23] and up to 50% of AML patients will relapse [23, 24]. On the contrary, CML treatment was revolutionized by the use of targeted tyrosine kinase inhibitors (TKIs), which have dramatically improved long-term survival [20, 25]. Particularly in AML, a prominent component of the disease is the recurrent mutations in spliceosome machinery and genome-wide aberrant splicing events [26]. As AS is an important part of normal hematopoiesis and necessary for cellular differentiation [27], the role of abnormal AS in leukemia progression and drug resistance has gained attention as several recent studies have highlighted recurrent splice factor mutations as important drivers of hematological malignancies [28-30].

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 PP2A-B56a subunit. Interestingly, NOCIVA is predominantly a nuclear protein. We also show that the expression of *NOCIVA* is elevated in cancer and that in myeloid cancers, such as AML and CML, high *NOCIVA* expression is a marker of a poor clinical outcome. Of particular clinical relevance, in CML high *NOCIVA* expression is associated with resistance to first generation TKI imantinib, but this effect is not seen with patients treated with second generation TKIs, dasatinib or nilotinib.

Methods

See supplemental information for more detailed 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. Details in Supplemental Methods.

Quantitative real-time PCR (RQ-PCR)

The standard curve analysis for amplification efficiency and the melting curve analysis for NOCIVA#1 and NOCIVA#2 RQ-PCRs are shown in Supplemental Figure 1. The primer and probe sequences used in this study for RQ-PCR analysis are listed in Supplemental Table 1. Details in Supplemental Methods.

Binding assay

Protein expression, purification and binding assays were performed as in [1]. Details in Supplemental Methods.

NOCIVA antibodies

Two NOCIVA specific antibodies were generated by immunizing rabbits against NOCIVA specific peptide NNKNTQEAFQVTS by BioGenes GmbH (Berlin, Germany). Details in Supplemental Methods.

Patient cohorts

AML study cohort

Detailed information for AML study cohort can be found in [31]. All 80 patients received regimens comprising anthracycline and high-dose cytarabine as induction therapy. 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 European LeukemiaNet (ELN) 2010 risk group classification [32] was used for risk stratification (Table S2).

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 (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).

CML study cohort2

This cohort consisted of 159 newly diagnosed CML patients from the UK-wide SPIRIT2 clinical trial [33]. 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 (Q_1 = 43, Q_3 = 63) and median follow-up time was 60 months (range 1–60 months).

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. Details in Supplemental Methods.

Data sharing statement

For original data, please contact jukwes@utu.fi.

Results

Identification of 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 expected from database searches (NCBI databases, Ensembl), one of the observed variants contained CIP2A exons 1-19, instead of 21 exons in the full length *CIP2A*. In addition, a novel CIP2A mRNA splice variant (named here as *NOCIVA*) with alternative exon inclusion was identified (Fig 1A). *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 inside the 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 polyadenylation signal (PAS) AATAAA and poly(A) tail (Fig 1B and Fig S2A).

As an indication that *NOCIVA* mRNA is created by AS, the spliced intron region is flanked by GT and AG dinucleotides (Fig S2B yellow, GU-AG intron). Further, based on *in silico* analysis with Human Splicing Finder (version 3.1, [34]), the junction site between *CIP2A* and *NOCIVA* contains exonic splicing silencer (ESS) matrices, especially Fas ESS and PESS-octamers. ESSs work by

inhibiting the splicing of pre-mRNA strands or promoting exon skipping. On the other hand, SpliceAid 2 [35] and SFmap (version 1.8, [36]) gave identical predictions for binding of YB-1 and SRp20 (SRSF3) splicing factors at the *NOCIVA* junction site (Fig 1C). Both of these splice factors have been shown to promote exon-inclusion during alternative splicing [37, 38]. Further, binding sites for many other splice factors, including Sam68, SLM2, SRp40 and multiple hnRNPs (including hnRNP K), were found in the near vicinity of the junction site (Fig 1C).

Validation PCR for full length *NOCIVA* mRNA expression was conducted in HeLa cell line with forward primer targeting *CIP2A* exon1 combined with various reverse primers targeting the *NOCIVA* specific 3' end of the mRNA (Fig 1D and Fig S3A). Additionally, validation PCR for *NOCIVA* expression was conducted in multiple cancer cell lines with primers specific for mRNA coding for the unique C-terminal portion of *NOCIVA* (Fig 1E and Fig S3B). The correct size bands from the gels were subsequently sequenced to confirm that the PCR product represented the *NOCIVA* mRNA product.

Together, these results identify *NOCIVA* as a novel, alternatively spliced *CIP2A* variant that is expressed in multiple cancer cell lines.

Characterization of NOCIVA protein

Interestingly, in *NOCIVA* mRNA, the 5´end of the *NOCIVA* specific intronic sequence is fused in coding frame with the preceding 3´end of the *CIP2A*

mRNA sequence. After 40 nucleotides, corresponding to 13 amino acids (aa) (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 aa 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 with any known protein sequence in the human proteome based on Blast homology search [39] (Fig S4A, BLASTP 2.8.1+, Database: Non-redundant protein sequences (nr)). Next, 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 neither full length CIP2A nor CIP2A protein fragments (Fig 2A, Fig S4B for NOCIVA ab #2 data). Importantly, the NOCIVA signal could be abolished by using blocking peptide (Fig 2A). To study spatial expression of endogenous NOCIVA, we performed immunofluorescence (IF)-staining in MDA-MB-231 breast cancer cells. Interestingly, whereas CIP2A resided predominantly in the cytoplasm as expected [3], endogenous NOCIVA positivity was clearly nuclear (Fig 2B and Fig S4C). Similar conclusion could be drawn from GFP fusion overexpression studies, in which NOCIVA-GFP colocalized with DAPI to nucleus (Fig 2C). With this approach also cytoplasmic NOCIVA-GFP was detected which was probably due to prominent colocalization of empty GFP to cytoplasm (Fig 2C). Hence, NOCIVA expresses a novel immunogenic peptide sequence, and constitutes a predominantly nuclear CIP2A variant protein.

To address NOCIVA protein functions, recombinant GST-NOCIVA (CIP2A 1-545+13 aa peptide) and GST-CIP2A 1-560 were compared (Fig S4D for Coomassie staining) in two functions critical for CIP2A-mediated PP2A modulation; protein homodimerization, and direct binding to the B56α subunit of PP2A [1]. Consistently with the location of the 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). Additionally, NOCIVA was competent to heterodimerize with CIP2A 1-560, albeit with lower affinity than that was seen with CIP2A 1-560 homodimers (Fig 3B). This can be explained as the CIP2A-NOCIVA fusion site partly overlaps with the aa region mediating CIP2A homodimerization [1](Fig 3C,D) and that when compared to CIP2A homodimers, in NOCIVA-CIP2A heterodimers some of the stabilizing interactions are lost (Fig 3E).

Together, these results identify NOCIVA as a novel nuclear CIP2A variant protein, that can heterodimerize with CIP2A and bind directly to the B56α tumor suppressor subunit of PP2A.

NOCIVA expression in normal and cancer cells

To evaluate the expression levels of *NOCIVA* mRNA in biological samples, and to compare them with *CIP2A,* we designed and validated (details in supplementary methods) two quantitative real time PCR (RQ-PCR) assays for both *NOCIVA* (NOCIVA#1 and #2 assays) and CIP2A (CIP2A e13 and e20 assays). If not otherwise indicated, NOCIVA#1 and CIP2A e20 were the

mainstay assays when refering to *NOCIVA* or *CIP2A* mRNA detection in this study.

First, *NOCIVA* and *CIP2A* expression was analyzed in a panel of normal human tissue cDNAs (Human MTC panel I & II, Clontech, cat no 636742 & 636743). Notably, *CIP2A* and *NOCIVA* RQ-PCR assays were optimized to yield similar amplification efficiency allowing 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 than *CIP2A,* expression profile across different tissues, including high expression in testis, was comparable to that of *CIP2A* (Fig. 4A and S5B). To identify tissues in which *CIP2A* AS to *NOCIVA* might be relatively more active we calculated the ratio between *NOCIVA* and *CIP2A* expression across different normal tissues. Although the absolute expression of *NOCIVA* was below 7% of *CIP2A* in all tissues, the leukocytes, kidney, and pancreas were the tissues that had the highest *NOCIVA/CIP2A* ratio (Fig 4B).

To address potential overexpression of *NOCIVA* in cancer, we first assessed *NOCIVA* mRNA expression between normal epidermal keratinocytes (NHEK, Ker), and patient-derived head and neck squamous cell carcinoma (HNSCC) cells in which *CIP2A* is overexpressed (Fig S5C)[3, 40]. Interestingly, also *NOCIVA* mRNA showed significantly elevated expression in HNSCC samples as compared to NHEK (Fig 4C, p=0.0001 by Student’s t-test).

Followed by the highest *NOCIVA/CIP2A* ratio in lymphoid cells (Fig. 4B), we next tested whether this preferential *NOCIVA* expression was found also from lymphoid cancer cells. Indeed, 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 3D). Encouraged by these results, we validated preferential *NOCIVA* gene expression from a panel of clinical 80 AML (BM) and 35 (peripheral blood) CML samples. Consistently with earlier results [21, 31], 96% of AML, and 94% CML, patients, respectively, expressed lower levels of *CIP2A* than normal BM controls pooled from 56 males and females (Fig. 4E,F). However, fully supporting active AS from *CIP2A* to *NOCIVA* in myeloid cancers, 77% of AML, and 65% CML samples displayed overexpression of *NOCIVA* (Fig 4E, F).

AML samples were additionally analyzed for mutual dependencies in gene expression levels between *NOCIVA* and the established AML markers Wilms' tumor 1 (WT1)[41] and ectopic viral integration site-1 (EVI1)[42]; and the PP2A inhibitor proteins *CIP2A, SET, ARPP19, TIPRL,* and *PME1* [31]. Based on Pearson pairwise correlation analysis, we found that *NOCIVA* expression levels significantly correlated with *PME1* (r=0.43, p=0.0002) and weakly but significantly with *ARPP19* (r=0.37, p=0.0014) and *SET* (r=0.30, p=0.0104), but not with other studied markers (Fig 4G).

These results show that *NOCIVA* has a similar expression pattern across normal human tissues to that of *CIP2A*. However, *NOCIVA* displays robust

overexpression in AML and CML in contrast to *CIP2A* underexpression from the same samples.

Clinical relevance of *NOCIVA* expression in diagnostic AML samples

The results above indicate that the myeloid leukemias AML and CML are malignancies in which active splicing of *CIP2A* to *NOCIVA* is particularly prominent. To understand potential clinical significance of this AS phenomenon, we next analyzed the prognostic significance of *NOCIVA* mRNA expression in 80 AML cases treated with intensive chemotherapy (AML study cohort, [31]). After dividing *NOCIVA* expression into high and low expression according to median (2.18, Q_1=1.14, Q_3=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). Very interestingly, low *CIP2A* (Fig 5B, p=0.073 by log-rank test) expression instead was a borderline significance predictor of longer OS indicating that active AS from *CIP2A* to *NOCIVA* is oncogenic in AML.

Additional analysis for the prognostic role of the studied genes for OS was performed by Cox's proportional multivariable hazard model, which included age at diagnosis and diagnostic mRNA expression levels of *CIP2A e13, CIP2A e20, SET, ARPP19, TIPRL, PME1, EVI1, WT1* and *NOCIVA.* After excluding the non-significant 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 as independent prognostic factors for OS. It

was notable that the hazard ratio for *NOCIVA* mRNA expression was even higher than for either *EVI1* expression or diagnosis age which in clinical practise are considered as strong predictors of AML outcome.

We also analysed the association with clinical characteristics and risk groups for the studied markers. The expression of *NOCIVA* or *CIP2A* did not show correlations to any of the clinical characteristics; age, gender, leukocyte or BM blast count, secondary leukemia or the presence/absence of a normal karyotype. In regard to genetic risk group associations, neither *NOCIVA* nor *CIP2A* expression levels showed an association with the ELN2010 risk groups (Fig 5D, $p>0.05$ by Kruskal-Wallis test). On the other hand, and as expected, *EVI1* mRNA expression at diagnosis was significantly different between the three risk groups, and its expression increased in accordance to the risk group ($p=0.005$ by Kruskal-Wallis test).

Together these data demonstrate a significant, but risk group independent, association between high *NOCIVA* expression 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 cohort1). Twenty patients received imatinib (1G TKI), and 14 dasatinib or nilotinib (2G TKI), as the 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 according to median (5.5, Q_1=0.20, Q_3=20.0), analysis revealed that high *NOCIVA* mRNA expression was associated with significantly shorter EFS (Fig 6A, p=0.024 by log-rank test).

Very 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 the patients treated with 2G TKI (Fig. 6C, p=0.429 by log-rank test). Time to Complete Molecular Response (CMR) analysis 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, no patient with high levels of *NOCIVA* mRNA at diagnosis achieved CMR. Again, among the patients treated with 2G TKI´s, no association was found between *NOCIVA* expression and CMR, indicating that 2G TKI therapy may overcome the adverse effect of high mRNA expression of NOCIVA.

These findings were then validated in the 159 patient independent CML study cohort2 from the SPIRIT2 clinical trial [33]. In this cohort, 81 patients had received imatinib and 78 dasatinib as the first line therapy. As seen in the CML study cohort1, also here high *NOCIVA* expression at diagnosis was associated with disease progression only among the imatinib-treated patients. Imatinib-treated patients who subsequently progressed to blast crisis had significantly higher expression of *NOCIVA* at diagnosis, than those patients who did not

409 progress (Fig 6E, p=0.04 by Mann Whitney u-test). No significant difference
410 was observed for those patients treated with dasatinib (Fig. 6E). Interestingly,
411 imatinib-treated patients with highest quartile *NOCIVA* expression at diagnosis
412 had significantly inferior freedom from progression (FFP) as compared to
413 patients with lower *NOCIVA* expression (Fig 6F, p=0.039 by log-rank test).
414 Consistent with results from CML study cohort1, no association between
415 *NOCIVA* expression and FFP was observed among the dasatinib-treated
416 patients (Fig 6G).
417
418 In conclusion, high NOCIVA mRNA expression assessed at CML CP diagnosis
419 is associated with an inferior EFS and FFP as well as lower rates of CMR
420 selectively among the imatinib-treated patients.
421

Discussion

Cancerous inhibitor of PP2A (CIP2A) is an oncoprotein with clinical relevance in a number of human cancers [2]. Furthermore, CIP2A is a very attractive therapeutic target as it is a direct inhibitor of tumor suppressor PP2A-B56a [1], and has low expression in normal tissues [3]. However, regardless of thorough profiling of CIP2A expression across human cancers, mRNA or protein variants of CIP2A remain uncharacterized. In NCBI The Nucleotide and The Protein database, there are four predictions for *CIP2A* splice variant as well as for protein variants of CIP2A. However, none of the predicted variants resemble *NOCIVA,* and neither their functional nor clinical relevance have been studied. This notion emphasizes both the novelty of the presented work as well as indicates the obvious need for experimental validation of also the predicted *CIP2A* isoforms to comprehensively understand CIP2A regulation and function in cancer.

One of the most interesting features of *NOCIVA* is that the intronic region spliced to nucleotide 1636 of *CIP2A*, the last nucleotide of the triplet the codes CIP2A aa 545, is in coding frame with the preceding *CIP2A* sequence and thus codes for a unique, immunogenic C-terminal 13 aa peptide tail (Fig 1B). NOCIVA can be considered as a novel human protein as no sequences homologous could be identified to the 13 aa NOCIVA tail in the human proteome. Strongly indicative of alternative cellular functions of CIP2A and NOCIVA, 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 to bind to B56a, indicating that it functions similar to CIP2A as a PP2A inhibitor protein. It is clear that further studies on the differential functional roles of NOCIVA and CIP2A are warranted. Unfortunately, during the project we failed in development of siRNA or CRISPR/Cas9 tools selectively suppressing *NOCIVA*, and therefore rather invested thoroughly in demonstrating clinical relevance of the discovery of *NOCIVA*. It is however clear, that thoroughly validated functional models are needed in the future to untangle the cellular role of *NOCIVA*.

AML and CML patient samples displayed clearly higher expression of *NOCIVA* mRNA over *CIP2A* suggesting for particularly active AS of *CIP2A* in myeloid malignancies. This is interesting as AML and CML are the only cancer types where CIP2A seems to be underexpressed at the mRNA level as compared to normal tissue [21, 22]. We thus postulate that upon leukemogenesis a splicing switch that creates *NOCIVA* from *CIP2A* is activated. As splicing is a highly complex event, also at the *NOCIVA* junction site, exonic splicing silencer sequences (ESS) as well as binding sites for hnRNPs and splice factor can be found (Fig 1C). ESSs work by inhibiting the splicing of pre-mRNA strands or promoting exon skipping by primarily recruiting hnRNP binding. A recent paper reported that the most significant Gene Ontology ‘Processes’ and ‘Networks’ changed in AML blasts compared to normal controls were related to transcription, mRNA processing, and stabilization [44]. They observed changes in the expression of 13 hnRNPs affecting mRNA processing, out of which hnRNP A1, A2B1, C are predicted to bind to the *NOCIVA* junction site. Additionally, the expression of hnRNP K [45], SRSF3 (SRp20) [46] and YB-1

[47] have been shown to be altered in AML, but also to contribute to leukemia progression. Interestingly, SRSF3 [37] and YB-1 [38] are additionally shown to specifically promote exon-inclusion during AS. Detailed analysis of the role of these splice factors in AS of *CIP2A* to *NOCIVA* will be required in the future to better understand regulation of NOCIVA in myeloid cancers.

Our data indicate that high *NOCIVA* mRNA expression associates with poor outcome both in AML and CML. In AML, *NOCIVA* expression was independent of the current genetic risk classification, suggesting that the evaluation of *NOCIVA* expression at diagnosis could provide clinically relevant additional predictive value. We also found that high *NOCIVA* mRNA expression assessed at CML CP diagnosis is associated with an inferior EFS and FFP as well as lower rates of CMR for imatinib treated CML patients. Hence, the data suggest that 2G TKI therapy is required to overcome the adverse effects of high *NOCIVA* expression and that together with other diagnostic biomarkers, detection of *NOCIVA* at CML diagnostic phase might help in treatment decisions between imatinib and 2G TKI.

In summary, this work describes discovery of a novel human gene and protein product with the characteristics of a clinically relevant PP2A inhibitor in myeloid malignancies.

Acknowledgements

Taina Kalevo-Mattila is acknowledged for superior technical assistance. We thank Maria D. Odero (University of Navarra, Pamplona, Spain) for constructive discussions and advice on the material and methodology related to AML cell lines and CIP2A expression in AML. We also 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 want to thank Otto Kauko and Juha Okkeri for constructive discussions and advice on metholodology in the beginning of this project. This work was supported by funding from The Sigrid Juselius Foundation, Turku Doctoral Program of Molecular Medicine, University of Turku faculty of medicine, Turku University Hospital ERVA (13283), The Päivikki and Sakari Sohlbergin Foundation, The Cancer Foundation Väre, The Finnish Concordia Fund and Business Finland TUTL (2445/31/2017).

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.

519 **Disclosure of Conflicts of Interest**

520 J.W. and E.M have patents pending for “A NOVEL CIP2A VARIANT AND USES
521 THEREOF” (PCT/FI2018/050844) and “METHOD FOR PREDICTING
522 RESPONSE TO TREATMENT WITH TYROSINE KINASE INHIBITORS AND
523 RELATED METHODS” (PCT/FI2020/050257). In the past 3 years, R.E.C and
524 C.L. declare relevant research support from Bristol Myers Squibb, and R.E.C.
525 declares non-relevant research support and honoraria from Pfizer and Novartis
526 and non-relevant honoraria from Jazz pharmaceuticals and Abbvie.

527

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: 437-450.

2 Khanna A, Pimanda JE, Westermarck J. Cancerous inhibitor of protein phosphatase 2A, an emerging human oncoprotein and a potential cancer therapy target. *Cancer Res* 2013; 73: 6548-6553.

3 Junttila MR, Puustinen P, Niemelä M, Ahola R, Arnold H, Böttzauw T *et al*. CIP2A inhibits PP2A in human malignancies. *Cell* 2007; 130: 51-62.

4 Khanna A, Pimanda JE. Clinical significance of Cancerous Inhibitor of Protein Phosphatase 2A (CIP2A) in human cancers. *Int J Cancer* 2015.

5 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: 5092-5100.

6 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: 182-197.

551 7 Khanna A, Böckelman C, Hemmes A, Junttila MR, Wiksten JP, Lundin M
552 *et al*. MYC-dependent regulation and prognostic role of CIP2A in gastric cancer.
553 *J Natl Cancer Inst* 2009; 101: 793-805.
554
555 8 Elgendy M, Ciro M, Hosseini A, Weiszmann J, Mazzarella L, Ferrari E *et*
556 *al*. Combination of Hypoglycemia and Metformin Impairs Tumor Metabolic
557 Plasticity and Growth by Modulating the PP2A-GSK3beta-MCL-1 Axis. *Cancer*
558 *Cell* 2019; 35: 798-815 e795.
559
560 9 Kauko O, Imanishi SY, Kulesskiy E, Yetukuri L, Laajala TD, Sharma M *et*
561 *al*. Phosphoproteome and drug-response effects mediated by the three protein
562 phosphatase 2A inhibitor proteins CIP2A, SET, and PME-1. *J Biol Chem* 2020;
563 295: 4194-4211.
564
565 10 Kauko O, O'Connor CM, Kulesskiy E, Sangodkar J, Aakula A, Izadmehr
566 S *et al*. PP2A inhibition is a druggable MEK inhibitor resistance mechanism in
567 KRAS-mutant lung cancer cells. *Science translational medicine* 2018; 10.
568
569 11 Myant K, Qiao X, Halonen T, Come C, Laine A, Janghorban M *et al*.
570 Serine 62-Phosphorylated MYC Associates with Nuclear Lamins and Its
571 Regulation by CIP2A Is Essential for Regenerative Proliferation. *Cell Rep* 2015;
572 12: 1019-1031.
573

574 12 Ventelä S, Côme C, Mäkelä JA, Hobbs RM, Mannermaa L, Kallajoki M *et*
575 *al*. CIP2A promotes proliferation of spermatogonial progenitor cells and
576 spermatogenesis in mice. *PLoS One* 2012; 7: e33209.
577
578 13 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of
579 alternative splicing complexity in the human transcriptome by high-throughput
580 sequencing. *Nat Genet* 2008; 40: 1413-1415.
581
582 14 Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant
583 RNA splicing in cancer; expression changes and driver mutations of splicing
584 factor genes. *Oncogene* 2016; 35: 2413-2427.
585
586 15 Climente-González H, Porta-Pardo E, Godzik A, Eyras E. The Functional
587 Impact of Alternative Splicing in Cancer. *Cell Rep* 2017; 20: 2215-2226.
588
589 16 Kahles A, Lehmann KV, Toussaint NC, Hüser M, Stark SG, Sachsenberg
590 T *et al*. Comprehensive Analysis of Alternative Splicing Across Tumors from
591 8,705 Patients. *Cancer Cell* 2018; 34: 211-224.e216.
592
593 17 Yang YT, Chiu YC, Kao CJ, Hou HA, Lin CC, Tsai CH *et al*. The
594 prognostic significance of global aberrant alternative splicing in patients with
595 myelodysplastic syndrome. *Blood Cancer J* 2018; 8: 78.
596
597 18 Adamia S, Haibe-Kains B, Pilarski PM, Bar-Natan M, Pevzner S, Avet-
598 Loiseau H *et al*. A genome-wide aberrant RNA splicing in patients with acute

599 myeloid leukemia identifies novel potential disease markers and therapeutic
600 targets. *Clin Cancer Res* 2014; 20: 1135-1145.
601
602 19 Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T
603 *et al*. Diagnosis and management of AML in adults: 2017 ELN
604 recommendations from an international expert panel. *Blood* 2017; 129: 424-
605 447.
606
607 20 Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley
608 JF *et al*. European LeukemiaNet recommendations for the management of
609 chronic myeloid leukemia: 2013. *Blood* 2013; 122: 872-884.
610
611 21 Lucas CM, Harris RJ, Giannoudis A, Copland M, Slupsky JR, Clark RE.
612 Cancerous inhibitor of PP2A (CIP2A) at diagnosis of chronic myeloid leukemia
613 is a critical determinant of disease progression. *Blood* 2011; 117: 6660-6668.
614
615 22 Barragán E, Chillón MC, Castelló-Cros R, Marcotegui N, Prieto MI,
616 Hoyos M *et al*. CIP2A high expression is a poor prognostic factor in normal
617 karyotype acute myeloid leukemia. *Haematologica* 2015; 100: e183-185.
618
619 23 Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl*
620 *J Med* 2015; 373: 1136-1152.
621
622 24 Petersdorf SH, Kopecky KJ, Slovak M, Willman C, Nevill T, Brandwein J
623 *et al*. A phase 3 study of gemtuzumab ozogamicin during induction and

624 postconsolidation therapy in younger patients with acute myeloid leukemia.
625 *Blood* 2013; 121: 4854-4860.

626

627 25 Cross NC, White HE, Colomer D, Ehrencrona H, Foroni L, Gottardi E *et*
628 *al*. Laboratory recommendations for scoring deep molecular responses
629 following treatment for chronic myeloid leukemia. *Leukemia* 2015; 29: 999-
630 1003.

631

632 26 de Necochea-Campion R, Shouse GP, Zhou Q, Mirshahidi S, Chen CS.
633 Aberrant splicing and drug resistance in AML. *J Hematol Oncol* 2016; 9: 85.

634

635 27 Grech G, Pollacco J, Portelli M, Sacco K, Baldacchino S, Grixti J *et al*.
636 Expression of different functional isoforms in haematopoiesis. *Int J Hematol*
637 2014; 99: 4-11.

638

639 28 Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL *et*
640 *al*. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations.
641 *Blood* 2015; 125: 1367-1376.

642

643 29 Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P,
644 Roberts ND *et al*. Genomic Classification and Prognosis in Acute Myeloid
645 Leukemia. *N Engl J Med* 2016; 374: 2209-2221.

646

647 30 Taskesen E, Havermans M, van Lom K, Sanders MA, van Norden Y,
648 Bindels E *et al*. Two splice-factor mutant leukemia subgroups uncovered at the

649 boundaries of MDS and AML using combined gene expression and DNA-
650 methylation profiling. *Blood* 2014; 123: 3327-3335.
651
652 31 Mäkelä E, Löyttyniemi E, Salmenniemi U, Kauko O, Varila T, Kairisto V
653 *et al*. Arpp19 Promotes Myc and Cip2a Expression and Associates with Patient
654 Relapse in Acute Myeloid Leukemia. *Cancers (Basel)* 2019; 11.
655
656 32 Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK
657 *et al*. Diagnosis and management of acute myeloid leukemia in adults:
658 recommendations from an international expert panel, on behalf of the European
659 LeukemiaNet. *Blood* 2010; 115: 453-474.
660
661 33 O'Brien S, Cork L, Bandiera V, Bescoby R, Foroni L, Alaily L *et al*.
662 SPIRIT 2: Final 5 year analysis of the UK National Cancer Research Institute
663 randomized study comparing imatinib with dasatinib in patients with newly
664 diagnosed chronic phase chronic myeloid leukemia., vol. 132. Blood, 2018, p
665 (Supplement 1): 457.
666
667 34 Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M,
668 Béroud C. Human Splicing Finder: an online bioinformatics tool to predict
669 splicing signals. *Nucleic Acids Res* 2009; 37: e67.
670
671 35 Piva F, Giulietti M, Burini AB, Principato G. SpliceAid 2: a database of
672 human splicing factors expression data and RNA target motifs. *Hum Mutat*
673 2012; 33: 81-85.

36 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: W281-285.

37 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: 507-519.

38 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: 8622-8636.

39 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: 3389-3402.

40 Ventelä S, Sittig E, Mannermaa L, Mäkelä JA, Kulmala J, Löyttyniemi E *et al*. CIP2A is an Oct4 target gene involved in head and neck squamous cell cancer oncogenicity and radioresistance. *Oncotarget* 2015; 6: 144-158.

41 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: 217-221.

699 42 Hinai AA, Valk PJ. Review: Aberrant EVI1 expression in acute myeloid
700 leukaemia. *Br J Haematol* 2016; 172: 870-878.
701
702 43 Anande G, Deshpande NP, Mareschal S, Batcha AMN, Hampton HR,
703 Herold T *et al*. RNA Splicing Alterations Induce a Cellular Stress Response
704 Associated with Poor Prognosis in Acute Myeloid Leukemia. *Clin Cancer Res*
705 2020; 26: 3597-3607.
706
707 44 Alanazi B, Munje CR, Rastogi N, Williamson AJK, Taylor S, Hole PS *et*
708 *al*. Integrated nuclear proteomics and transcriptomics identifies S100A4 as a
709 therapeutic target in acute myeloid leukemia. *Leukemia* 2020; 34: 427-440.
710
711 45 Gallardo M, Lee HJ, Zhang X, Bueso-Ramos C, Pageon LR, McArthur M
712 *et al*. hnRNP K Is a Haploinsufficient Tumor Suppressor that Regulates
713 Proliferation and Differentiation Programs in Hematologic Malignancies. *Cancer*
714 *Cell* 2015; 28: 486-499.
715
716 46 Liu J, Huang B, Xiao Y, Xiong HM, Li J, Feng DQ *et al*. Aberrant
717 expression of splicing factors in newly diagnosed acute myeloid leukemia.
718 *Onkologie* 2012; 35: 335-340.
719
720 47 Liu S, Marneth AE, Alexe G, Walker SR, Gandler HI, Ye DQ *et al*. The
721 kinases IKBKE and TBK1 regulate MYC-dependent survival pathways through
722 YB-1 in AML and are targets for therapy. *Blood Adv* 2018; 2: 3428-3442.
723

724

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 number 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* stands for 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 bold and 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 and SFmap. At the *NOCIVA* junction site there are binding motifs for YB-1, SRp20 (SRSF3) and SRp40 splicing factors. **D)** Confirmation PCR to validate the *NOCIVA* specific full-length mRNA sequence expression from the Hela cell line. The forward primer for all lanes was the same *CIP2A* exon1 targeting, the reverse primer was *NOCIVA* specific and either R1, R2 or R3. Arrows indicating the right product. Resulting bands were extracted and the presence of specific *NOCIVA* cDNA (mRNA) was confirmed by DNA sequencing. **E)** Confirmation PCR to validate the *NOCIVA* specific mRNA sequence expression from several cell

lines with *CIP2A* exon13 targeting forward primer and the *NOCIVA* specific reverse primer. NTC stands for non-template control.

Figure 2. Characterization of NOCIVA protein in the cells

A) NOCIVA specific antibody detects correct size (appr. 90 kDa) recombinant GST-NOCIVA protein, but not recombinant CIP2A fragments. 1ug of each protein was loaded onto each gel. The signal is blocked with a NOCIVA-blocking peptide. Full length CIP2A comprises of 905 aa. **B)** Endogenous NOCIVA and CIP2A IF staining with NOCIVA specific antibody in MDA-MB-231 cells. NOCIVA green, CIP2A red, nucleus blue. Endogenous NOCIVA localizes mostly in the nucleus whereas full length CIP2A is mainly cytoplasmic. Scale bar 10 µm. **C)** NOCIVA-GFP overexpression in MDA-MB-231 cells. As seen with endogenous NOCIVA IF staining, GFP tagged NOCIVA also translocates to the nucleus. NOCIVA-GFP green, nucleus blue. Scale bar 10 µm.

Figure 3. Characterization of NOCIVA protein

A) *In vitro* GST-pulldown assay for interaction between PP2A B56α and GST-CIP2A(1–560), and GST-NOCIVA. Equal molar amounts of GST, GST-CIP2A(1–560) and GST-NOCIVA proteins were incubated with B56α for 1 h at 37°C before pulldown. Representative images from three experiments are shown. The accompanying 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. Equal molar amounts of GST, GST-CIP2A(1–560) and GST-NOCIVA proteins were incubated with CIP2A(1–560)-V5 fragment for 1 h at 37°C before pulldown. Samples were analyzed by Western blot using V5 and GST antibodies. Representative images from three experiments are shown. The graph shows relative dimerization efficiency of GST-NOCIVA as compared to GST-CIP2A (1–560), quantified as a ratio between CIP2A(1–560)-V5 and GST-CIP2A(1–560) in pulldown sample. Each bar is mean ± SD from three independent experiments; p = 0.017 by one sample t-test. **C)** CIP2A exists as obligatory dimer. Crystal structure of CIP2A(1-560) (PDB: 5UFL) shown as surface representation, with indicated dimer interface. Individual monomers in cyan and green. **D)** Zoom into dimer interface area modified in NOCIVA (modified part is in red line). Differences in NOCIVA (residues 546-560), in contrast to CIP2A, are mapped on CIP2A´s surface and shown in purple-blue and as transparent surface representation. Same orientation like in E. E-F were generated in Pymol. **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. In CIP2A-NOCIVA dimer stabilizing hydrogen bonds and salt bridges between S519-N553-R557-D520-Y556 and Q559-E560 are lost as compared to CIP2A-CIP2A dimer. Image was done using UCSF Chimera1.14. 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* mRNA expression in patient derived normal human epidermal keratinocytes (NHEK, Ker) and squamous cell carcinoma (SCC) cells. mRNA expression levels in Ker 45B cells was defined as value 1. p = 0.0001 by Student's t-test**. C)** *NOCIVA*/*CIP2A* mRNA expression in normal cells, indicating leucocytes as a tissue for further study. **D)** *NOCIVA* and *CIP2A* mRNA expression in AML, CML and solid cancer cell lines. AML: Acute myeloid leukemia; CML: Chronic myeloid leukemia. mRNA expression levels in Hela cells were defined as value 1. **E)** and **F)** 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. In the majority of cases, *CIP2A* mRNA expression in AML (E) and CML (F) patient samples was lower than in normal control samples, whereas *NOCIVA* was higher than in normal samples. **G)** 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) and weakly with *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 is associated with significantly worse overall survival in AML patients.

A) Kaplan–Meier survival curve for overall survival (OS) by *NOCIVA* gene expression in the AML patient cohort, stratified according to the median expression (see text). Higher *NOCIVA* expression is associated with shorter OS; p = 0.022 by log-rank test. **B)** No significant association was found between *CIP2A* gene expression level and OS (p = 0.073 by log-rank test), though there was a trend towards lower *CIP2A* expression being an indicator of shorter OS. **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 with risk groups in AML patient cohort by Kruskal-Wallis test. *NOCIVA* expression is a risk group independent prognostic marker in AML. As expected, *EVI1* mRNA expression at diagnosis was significantly different between the three risk groups and its expression increased in relation to the risk group; p = 0.005 by Kruskal-Wallis test. Group 1=favourable, 2=intermediate, 3=adverse. The ELN2010 genetic risk classification was used for risk stratification (see Supplementary Table 2 for detailed information).

Figure 6. High *NOCIVA* expression is associated with significantly worse outcome, including disease progression, in CML patients treated with imatinib.

A) Kaplan–Meier survival curve for event free survival (EFS) by *NOCIVA* mRNA expression in CML patient cohort1. 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 patient 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. Q_4 = 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. Q_4 = highest quartile of *NOCIVA* mRNA expression. 1G = first generation, 2G = second generation, TKI = tyrosine-kinase inhibitor. The median level of NOCIVA mRNA expression was used to define the high and low groups in each panel.

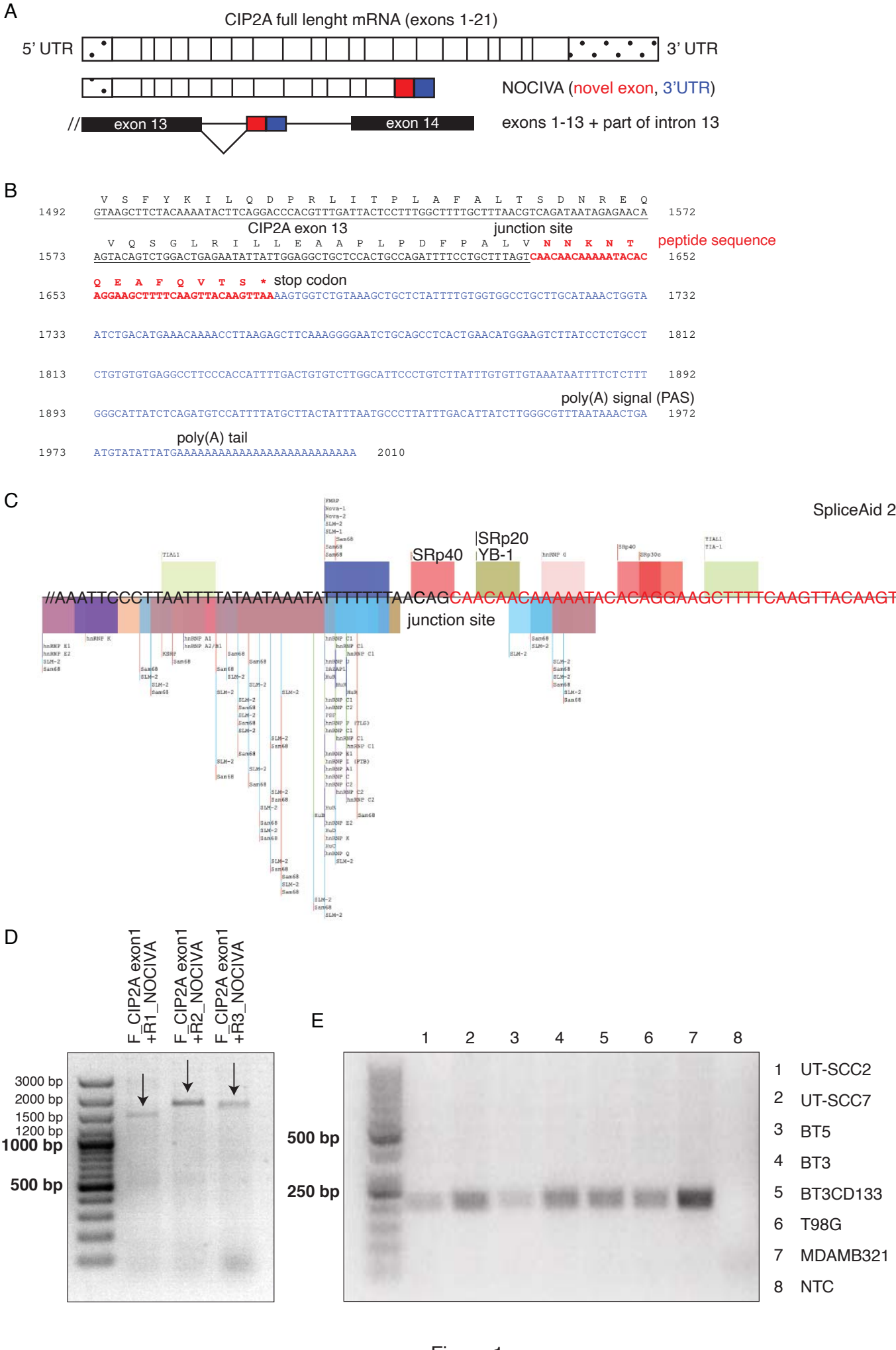

A
CIP2A full lenght mRNA (exons 1-21)
5' UTR
3' UTR
NOCIVA (novel exon, 3'UTR)
exon 13
exon 14
exons 1-13 + part of intron 13
B
CIP2A exon 13
junction site
peptide sequence
stop codon
poly(A) signal (PAS)
poly(A) tail
C
SpliceAid 2
SRp40
SRp20
YB-1
junction site
D
F_CIP2A exon1 +R1_NOCIVA
F_CIP2A exon1 +R2_NOCIVA
F_CIP2A exon1 +R3_NOCIVA
3000 bp
2000 bp
1500 bp
1200 bp
1000 bp
500 bp
E
1 2 3 4 5 6 7 8
500 bp
250 bp
1 UT-SCC2
2 UT-SCC7
3 BT5
4 BT3
5 BT3CD133
6 T98G
7 MDAMB321
8 NTC

Figure 1

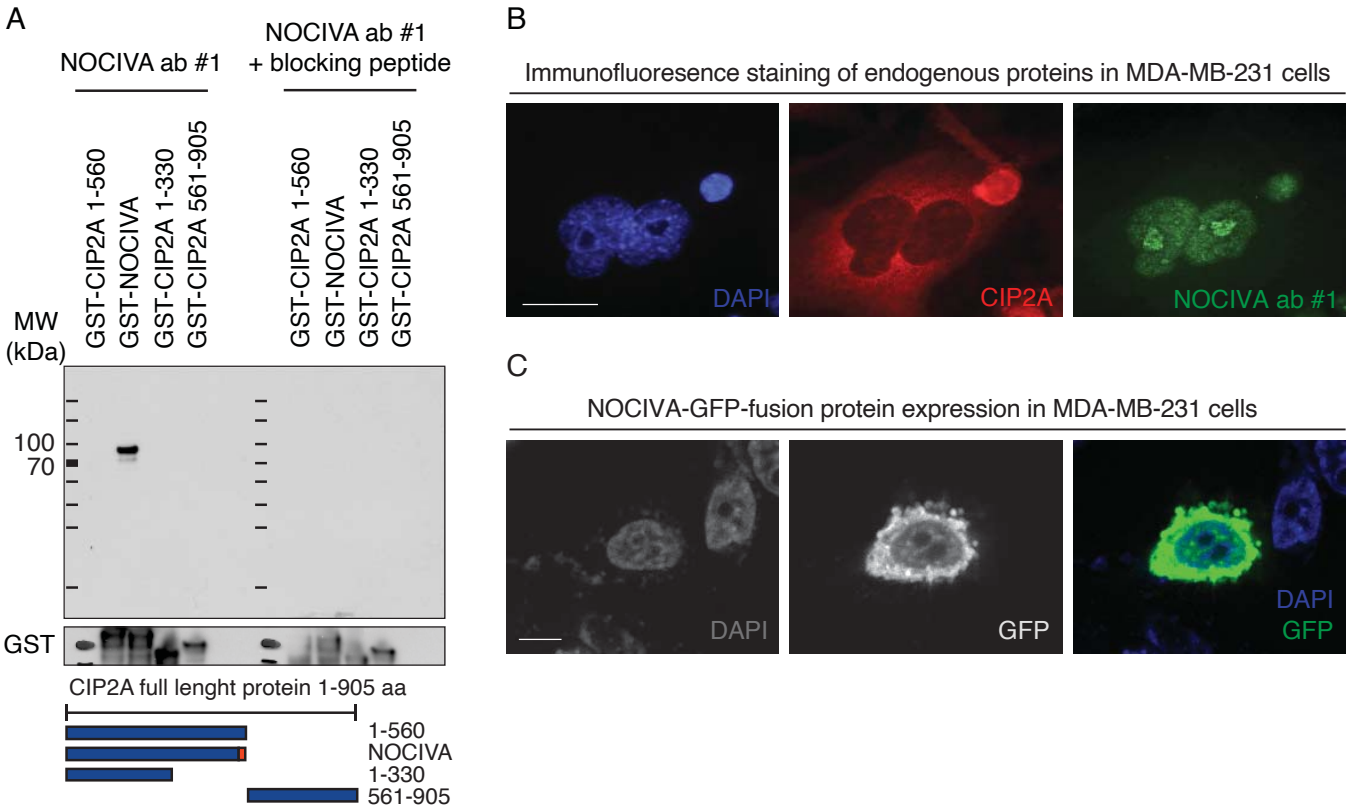
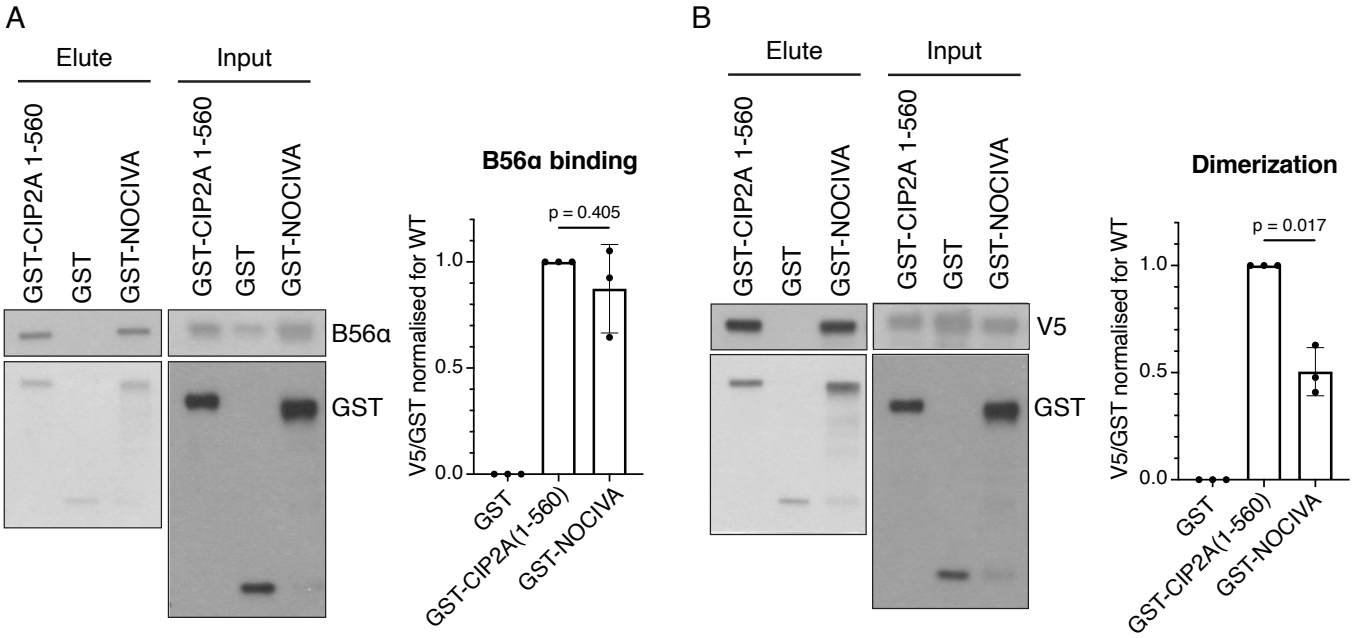
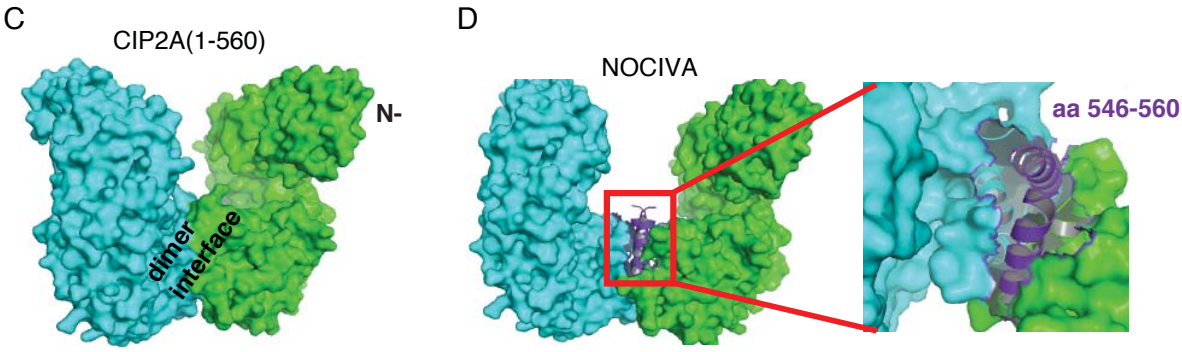
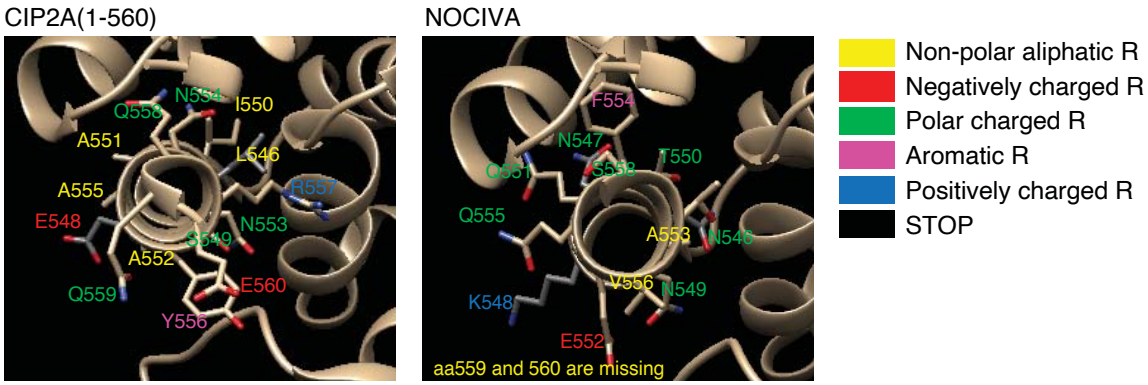

A
NOCIVA ab #1
NOCIVA ab #1
+ blocking peptide
GST-CIP2A 1-560
GST-NOCIVA
GST-CIP2A 1-330
GST-CIP2A 561-905
GST-CIP2A 1-560
GST-NOCIVA
GST-CIP2A 1-330
GST-CIP2A 561-905
MW
(kDa)
100
70
GST
CIP2A full lenght protein 1-905 aa
1-560
NOCIVA
1-330
561-905
B
Immunofluoresence staining of endogenous proteins in MDA-MB-231 cells
DAPI
CIP2A
NOCIVA ab #1
C
NOCIVA-GFP-fusion protein expression in MDA-MB-231 cells
DAPI
GFP
DAPI
GFP

Figure 2





E



aa	546	547	548	549	550	551	552	553	554	555	556	557	558	559
CIP2A	L	G	E	S	I	A	A	N	N	A	Y	R	Q	Q
NOCIVA	N	N	K	N	T	Q	E	A	F	Q	V	T	S	Stop

Figure 3

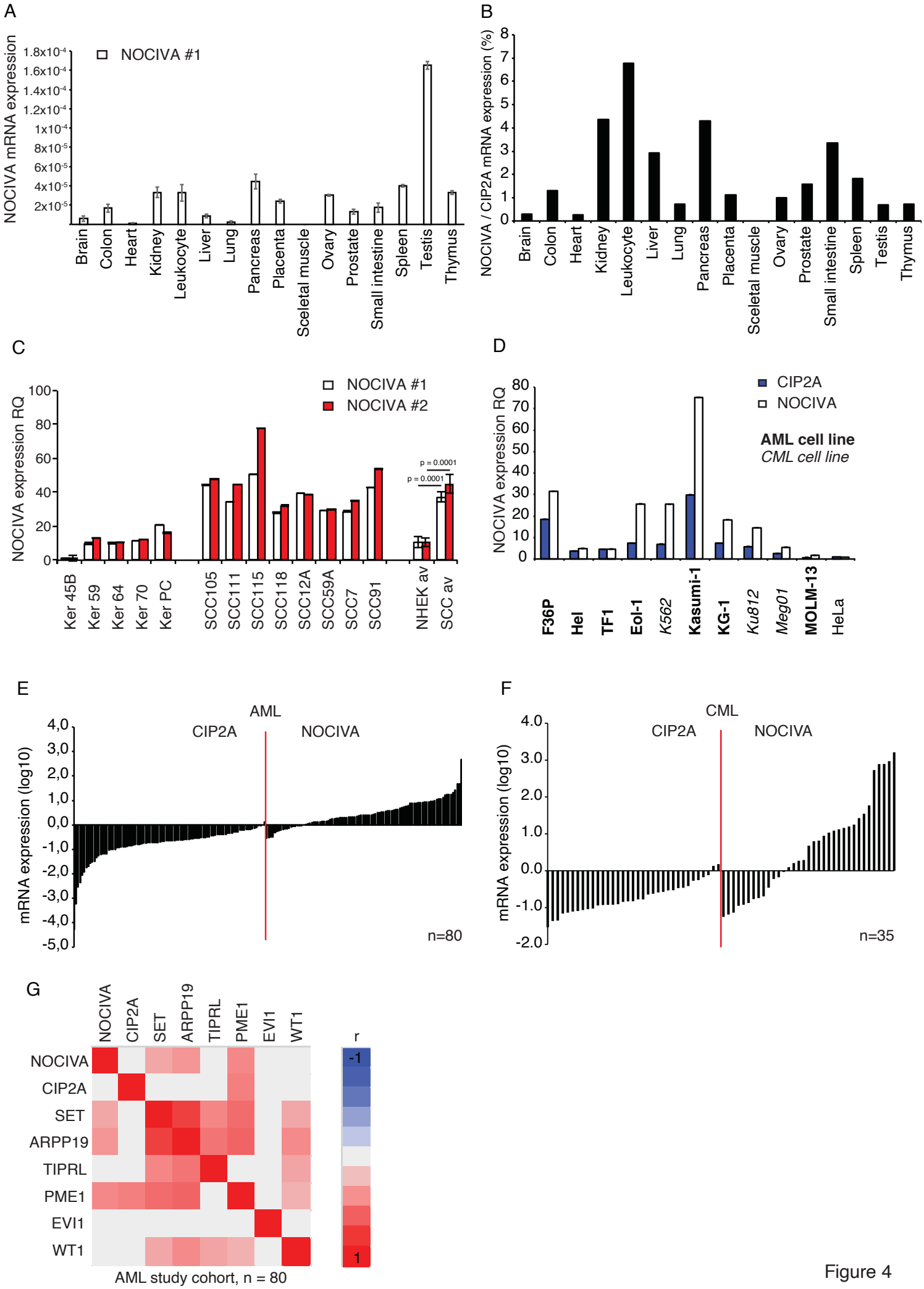
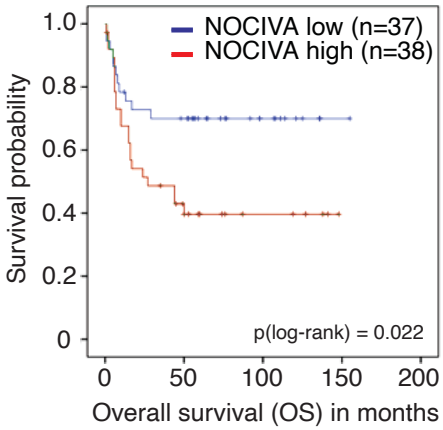
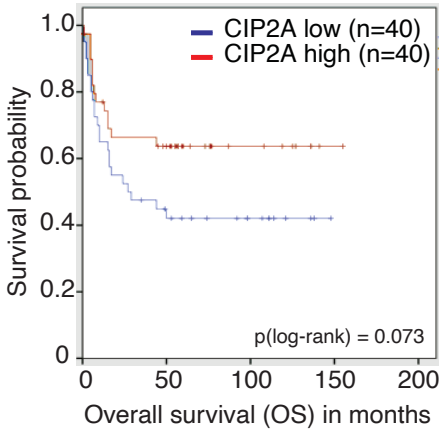

A
NOCIVA mRNA expression
NOCIVA #1
Brain
Colon
Heart
Kidney
Leukocyte
Liver
Lung
Pancreas
Placenta
Sceletal muscle
Ovary
Prostate
Small intestine
Spleen
Testis
Thymus
B
NOCIVA / CIP2A mRNA expression (%)
C
NOCIVA expression RQ
NOCIVA #1
NOCIVA #2
p = 0.0001
p = 0.0001
Ker 45B
Ker 59
Ker 64
Ker 70
Ker PC
SCC105
SCC111
SCC115
SCC118
SCC12A
SCC59A
SCC7
SCC91
NHEK av
SCC av
D
NOCIVA expression RQ
CIP2A
NOCIVA
AML cell line
CML cell line
F36P
Hel
TF1
Eol-1
K562
Kasumi-1
KG-1
Ku812
Meg01
MOLM-13
HeLa
E
AML
CIP2A
NOCIVA
mRNA expression (log10)
n=80
F
CML
CIP2A
NOCIVA
mRNA expression (log10)
n=35
G
NOCIVA
CIP2A
SET
ARPP19
TIPRL
PME1
EVI1
WT1
r
AML study cohort, n = 80

Figure 4

A



B



C

Parameter	Overall survival	
	Hazard Ratio	p
Diagnosis age	1.07	0.0013
EVI1	1.27	0.0004
NOCIVA	1.51	0.0205

D

Variable	Difference between groups 1-2-3	1 vs 2	1 vs 3	2 vs 3
WT1	+/- (0.077)	-	-	-
EVI	+ (0.005)	-	-	-
CIP2A	-	-	-	-
NOCIVA	-	-	-	+/- (0.074)

Figure 5

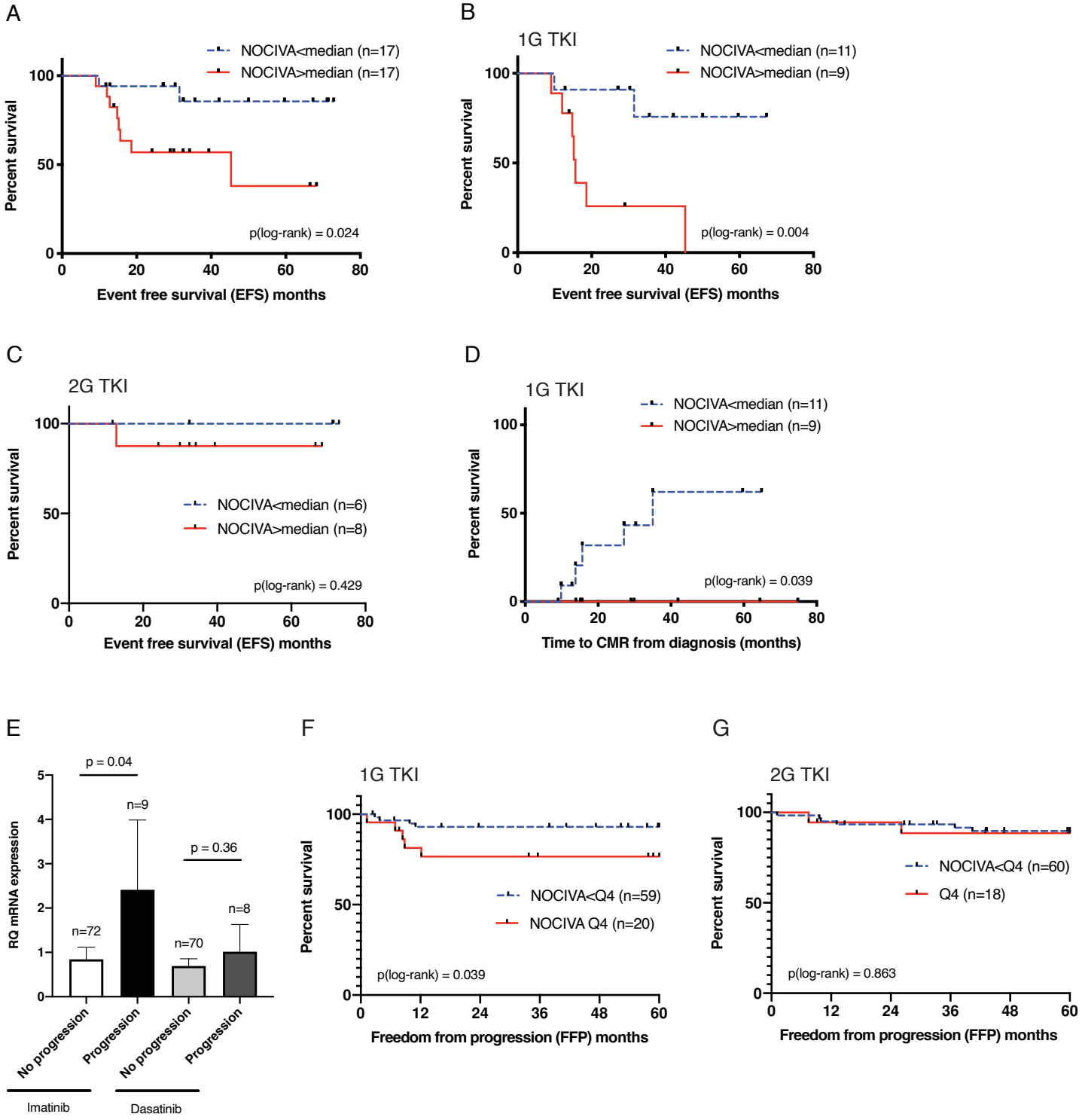
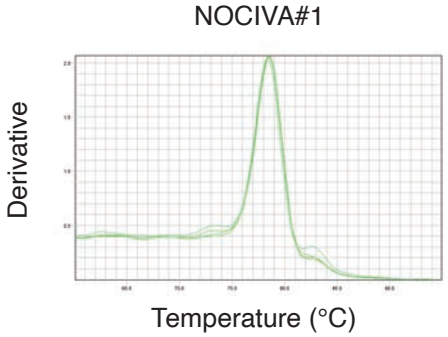
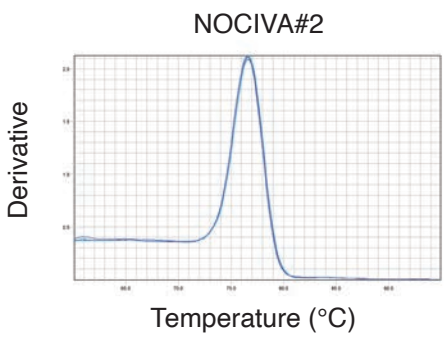

A
NOCIVA<median (n=17)
NOCIVA>median (n=17)
Percent survival
p(log-rank) = 0.024
Event free survival (EFS) months
B
1G TKI
NOCIVA<median (n=11)
NOCIVA>median (n=9)
Percent survival
p(log-rank) = 0.004
Event free survival (EFS) months
C
2G TKI
NOCIVA<median (n=6)
NOCIVA>median (n=8)
Percent survival
p(log-rank) = 0.429
Event free survival (EFS) months
D
1G TKI
NOCIVA<median (n=11)
NOCIVA>median (n=9)
Percent survival
p(log-rank) = 0.039
Time to CMR from diagnosis (months)
E
p = 0.04
p = 0.36
n=72
n=9
n=70
n=8
RQ mRNA expression
No progression
Progression
No progression
Progression
Imatinib
Dasatinib
F
1G TKI
NOCIVA<Q4 (n=59)
NOCIVA Q4 (n=20)
Percent survival
p(log-rank) = 0.039
Freedom from progression (FFP) months
G
2G TKI
NOCIVA<Q4 (n=60)
Q4 (n=18)
Percent survival
p(log-rank) = 0.863
Freedom from progression (FFP) months

Figure 6

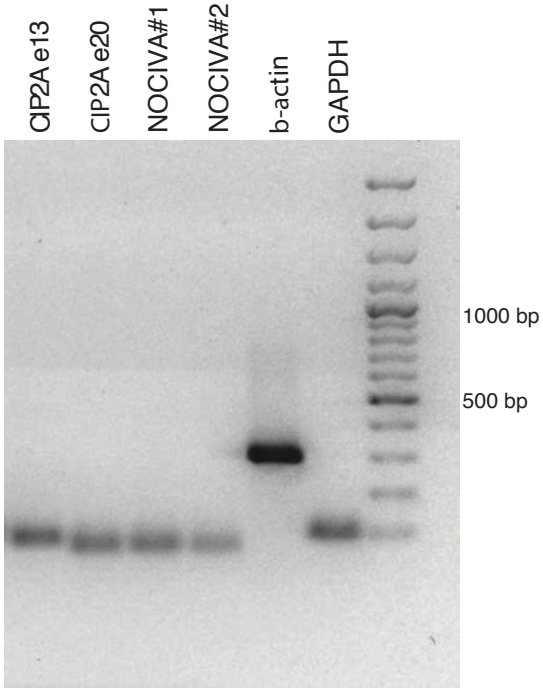
A



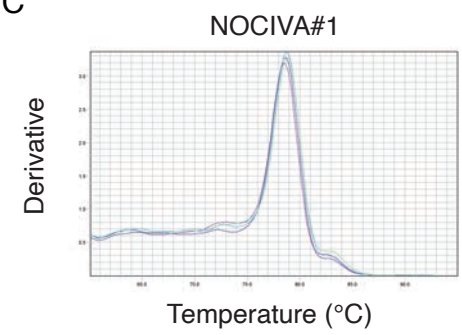
B



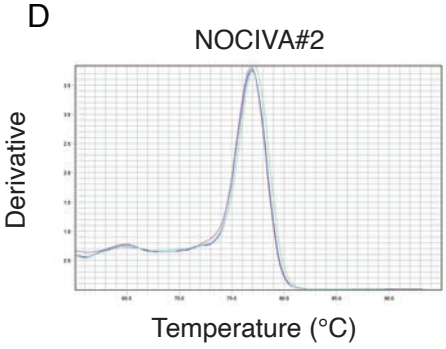
J



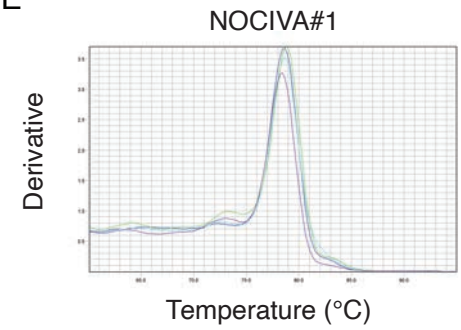
C



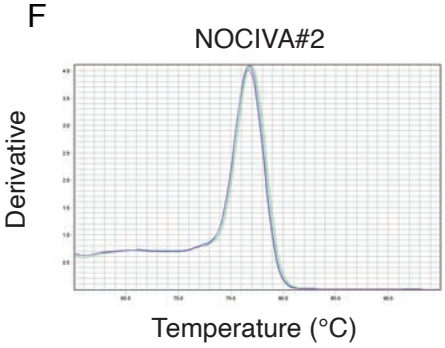
D



E

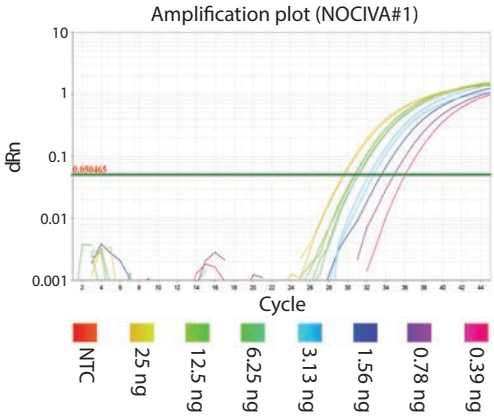


F

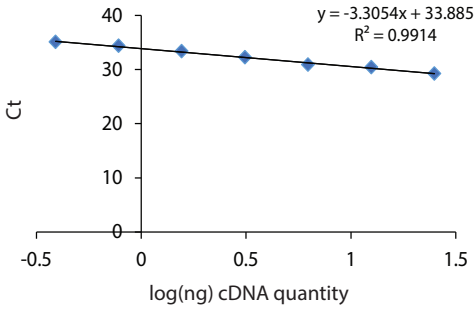


A-b=Hela
C-D=#67
E-F=HBM
G=Hela

G

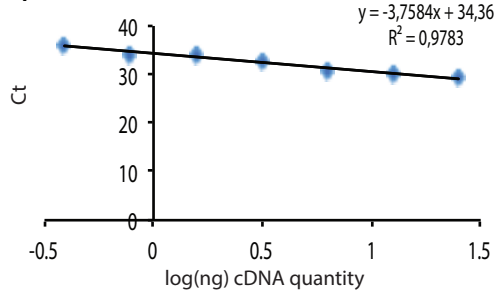


H



H=NOCIVA#1, #66 sample, E=101 %

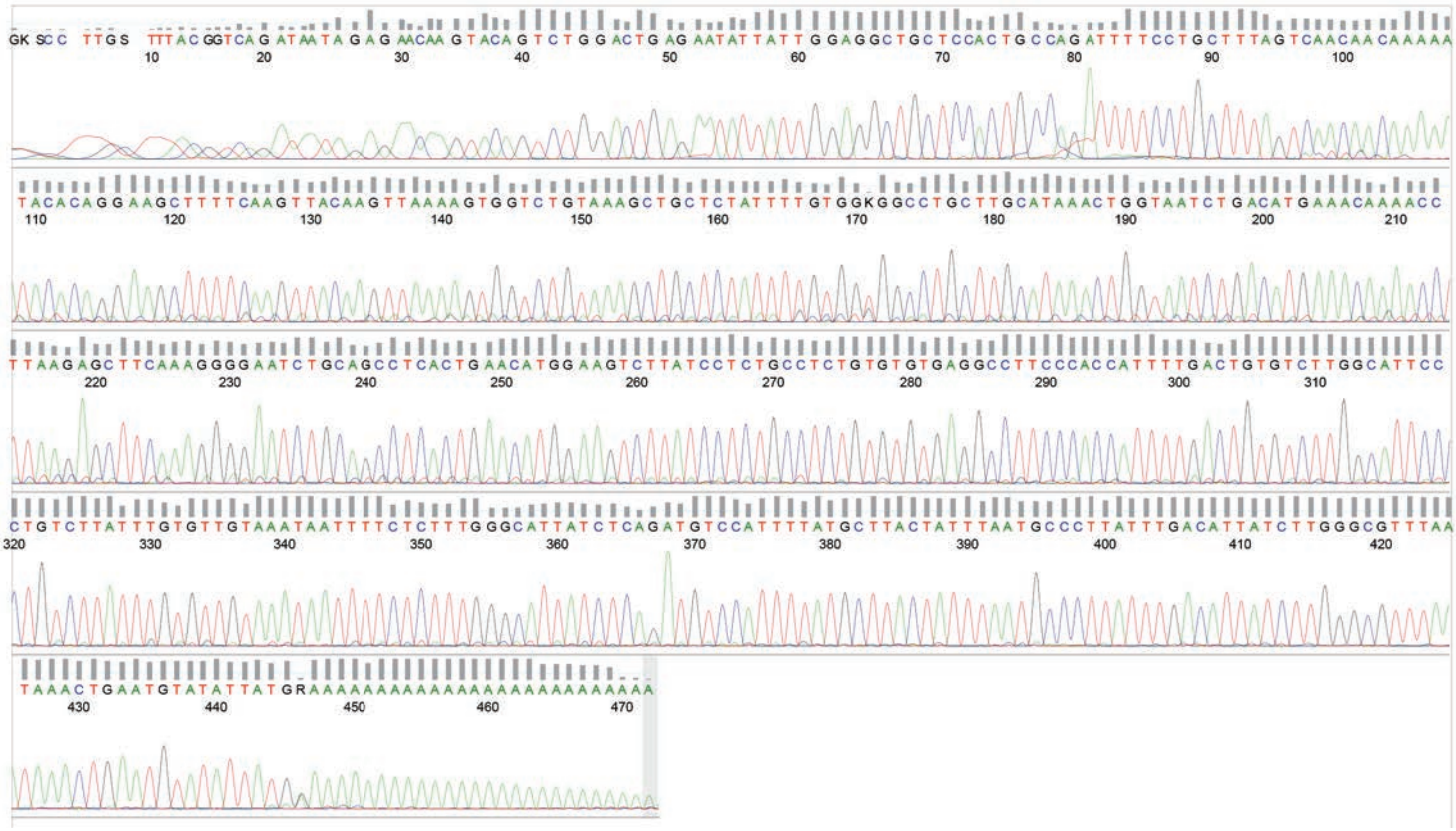
I



I=NOCIVA#2, #66 sample, E=85 %

Supplemental Figure 1

A



B

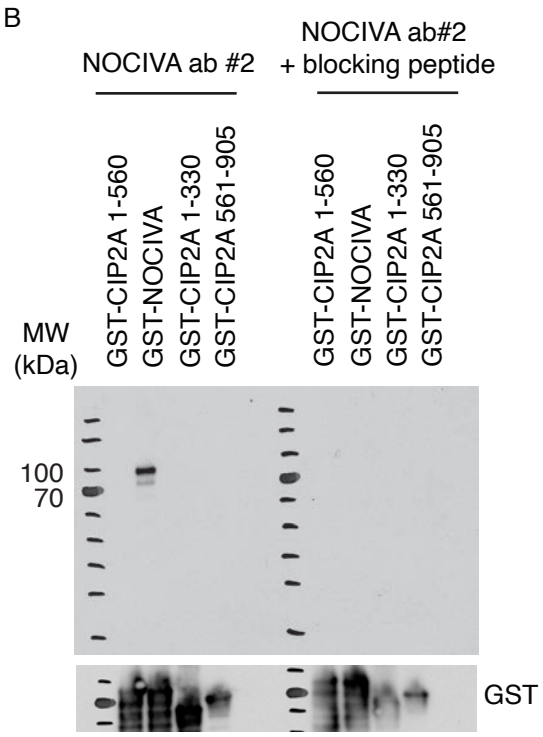
//ATTTATTTAG**GACCCACGTTTGATTACTCCTTTGGCTTTTGCTTTAACGTCAGATAATAGAGAACAAGTACAGTCTGGACTGAGAAT**
ATTATTGGAGGCTGCTCCACTGCCAGATTTTCCTGCTTTAGTGTGAGTTGTAATCATTTGCAGTGTATCTCTTGTGGTGTTGGAATACC
CTGCTATGTCTCTTTAGTCCTCAGTATAGATACACAAGCAAATTAGTGAATTGAAGACTTCATTGTAATAGTAATTTTTCAAGTGATTT
CTAAAAATGTTTGATTCTTATAAATGAAAAGTTCAGATGTATACTCACTCAGGGAAAAGCTAGATCCAGGCAGTTAAATGTTCTTCAGG
TACTATCCCTCTATGTCTCTAGCCACTTCTTTTTAATGATAACTTCTGGACACTCTTAAGCCTACATCCTCCCAGTTTAGTTCCTCCAT
AGAAAGAGCTTTTCCTTCCTGACTGTTTAAGGAAAAGGTCTAAGCATTAGTCATTGACCAGGCCTGGGTTATCCATGAGAACGATGTGG
ACTAAAAGTTAGTTGGGATAGGAGAATAGGAGAGTTCCTAAAGAAAAAAATCTGAACTTTTATCAGAAAGTGGAAAAATGACAACCAGG
CAAAATTTTCTACTATACTCGTTACTTTCAAGCAATAAGAATATTTGCAATTATGTTTGAAAAAAAATTCAGATTTCAGGGCTCTCTGA
ACTAAATGTCTATTAGGAGTCTGGTTTAAAATTTTTTTAAGTCATTTATATGTATGAATTTGTAAAAATGGTTTGAATAATTTGAAGTA
GTTGATAAACAGGTAAGATATTCCCTGTATACCGGGATCAGCAAACTGGCCAGTGGACAGGCATCTGTTTTTGTAAATAGTTTTATTGG
AATATGGCCACACCCACTCATTTTTATACTCTGTCTATGTTGCTTTCACTCTATAAAAGCAGAATTGAGTAATTGTGACAGAGACCAGA
TAGGTCCTTGAGTCTAAAATACTTACTATCTTGACCTATAAGAAAAAGTGCTGGTCCATGCTATATGCTATTCACTGATTTAGCGATTC
CACTTTTCAGAATTAGCAATTTAGGTCATCATACCTTAAGAGTTAAAGTATGGGATGTAAGACCTAGAGAACATATAACCAAAATTCCC
TTAATTTTATAATAAATATTTTTTTAAC**AG**CAACAACAAAAATACACAGGAAGCTTTTCAAGTTACAAGTTAA**AAGTGGTCTGTAAAGC**
TGCTCTATTTTGTGATGGCCTGCTTGCATAAACTGGTAATCTGACATGAAACAAAACCTTAAGAGCTTCAAAGGGGAATCTGCAGCCTC
ACTGAACATGGAAGTCTTATCCTCTGCCTCTGTGTGTGAGGCCTTCCCACCATTTTGACTGTGTCTTGGCATTCCCTGTCTTATTTGTG
TTGTAAATAATTTTCTCTTTGGGCATTATCTCAGATGTCCATTTTATGCTTACTATTTAATGCCCTTATTTGACATTATCTTGGGCGTT
TAATAAACTGAATGTATATTATGGATGCATTGGTAGGTAGGGTTTTTGTGTGTGTGTATTTACTTGTTTGTTTTCGAGACAGGGTCTCA
CTCTATCACCCAGACTGGGGTGCAGTGGTGTGATCATGACTCACTGCAGCTTCGACCTCTGGGGCTCAAGTGATATTCCCAACTCAACC
CCCTGAGTAGCGGGGACCACAGGCTTGTGCCACCACAACTGCTTAATTAATTTATTTTTTTATAGAGGGGTCTCACCAGGTTGCCCAGG
CTTGTCTCGAACTCTTGGACTCAAGTGATCCTCCCAAAGTGCTGCGGTTACAGGTGTGAGCCACCATGCCCAGCCTATAATAAATTAGG
CTGAGTGCCTGCCATCTTGCAGTTTGCACTTTAGTTAAGAGATAAGTGAAAAAAATATTTTGTTTTGTGGTAGAAAGCGTGTTCCAGAT
AAAGGAACAGCTCATATGAAGGCCTGAGGTAGGAGAAAGAGAGGATTTTAAAGGGTTAGAATGGAAGTCAGGTTGGTGAAAAGTATGTA
CTGAGTGAGAAAGGAAAAGGACTAAGGACGTAAAGGCCAGGTGATATCACCCATGATAAGAACATTATTCTGACCAGTACCAATAACTT
TGTACTTATTACAGATAATCCACACTCATAAACTAGTCTTACAATGCCTCCAAACCAGAGGTGAAGCTGTCATTATAAAAGGTTAGAAA
CAGAAAAGTGGGTGCTCTGGAAAATTCAGAATATTTCCCAAGACTTCATTTCATCATGGTCTCCAAGATTATTTGAAGACTTCTATTCA
AAAAAAGGACCCATTCTTAAAAATTCCCTAATTCTGCACTAAGTCTGCCTTTTCTTCTATTTTCGTATAATTTTTTTTTCCTTTCTTTG
ACTCCCATTCAG**ACTTGGAGAAAGTATAGCAGCAAACAATGCCTATAGACAACAGGAAACAGAACATATACCCAGAAAAATGCCCTGGC**
AATCATCAAATCACAGTTTTCCAACATCAATAAAGTGTTTAACTCCTCATTTGAAAGATGGTGTTCCTGGATTGAATATTGAAGAATTA
ATAGAGAAACTTCAGTCTGGAATGGTGGTGAGTGAAAAAATAGCAATTATAACCTGGTACATATTAGCTATATGTCAGTCCCAATTATA
ACGTGTATCTAAACACTGTCAGGTACAGATTTCACTATTGAAACCTTAATATAAGAAATAACAATATTCTGGAAAATTTGGAAATAAGA
AAAATCAGTCAAAATTTCAACATTTTGCCGGGTGCGGTGGCTT//

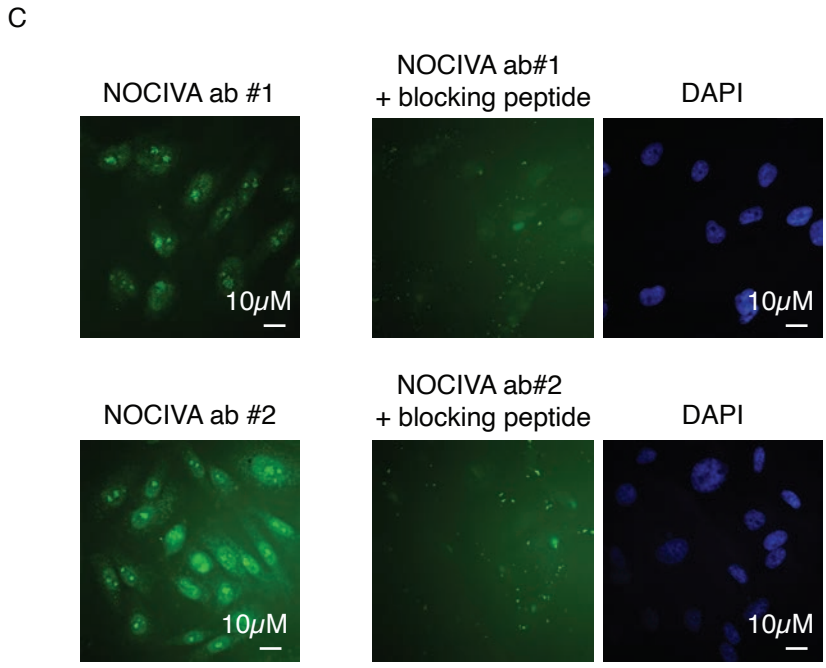
Supplemental Figure 2

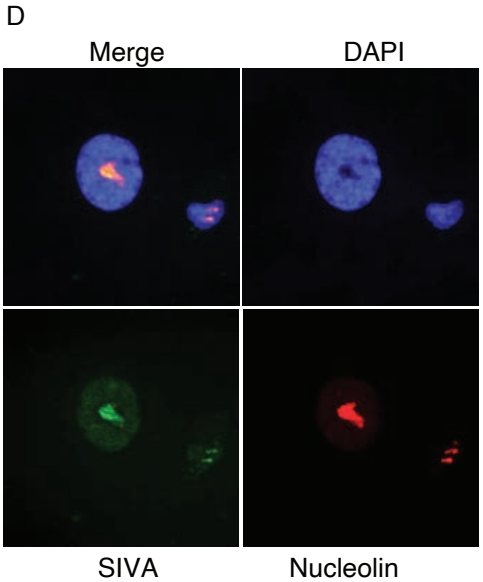
A

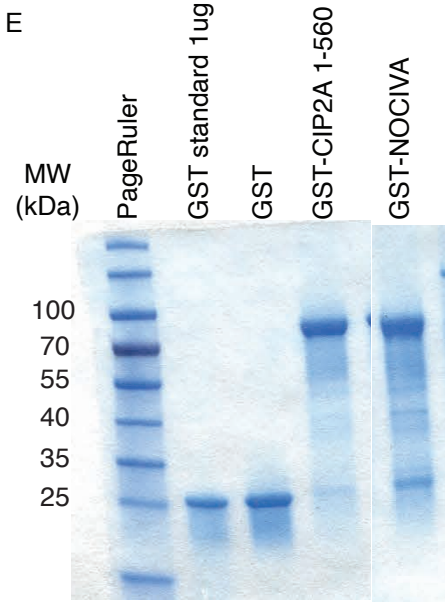
Sequences producing significant alignments with NOCIVA (NNKNTQEAFQVTS) in BLASTP search for Homo sapiens:

		Max Score	Query cover	E-value	Identity
NP_001001715.2	FERM, RhoGEF and pleckstrin domain-containing ...	23.5	61%	17	75%
BAF83484.1	unnamed protein product [Homo sapiens]	23.5	69%	17	78%
NP_005757.1	FERM, ARHGEF and pleckstrin domain-containing pro...	23.5	84%	17	75%
NP_001273768.1	FERM, ARHGEF and pleckstrin domain-containing ...	23.5	84%	17	75%
AAH71592.1	FARP1 protein [Homo sapiens]	23.5	84%	17	75%
EAW78561.1	hCG1786642, isoform CRA_a [Homo sapiens]	22.7	61%	36	75%
AAH64971.1	IQGAP1 protein [Homo sapiens]	22.7	53%	36	86%
BAG65182.1	unnamed protein product [Homo sapiens]	22.7	53%	36	86%
AAI39732.1	IQ motif containing GTPase activating protein 1 [H...	22.7	53%	36	86%
NP_003861.1	ras GTPase-activating-like protein IQGAP1 [Homo s...	22.7	53%	36	86%



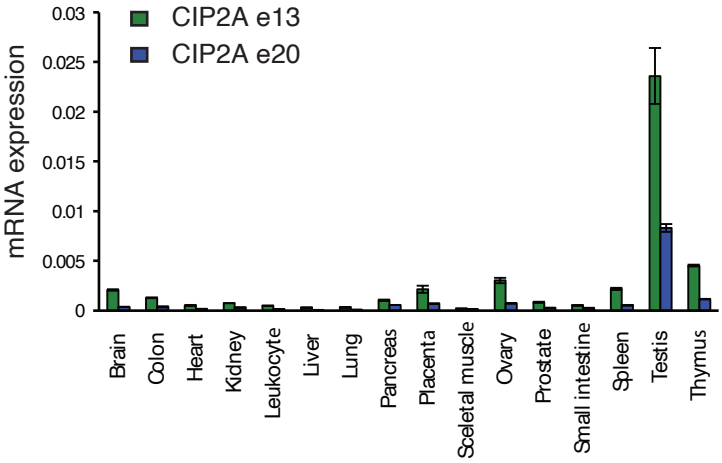




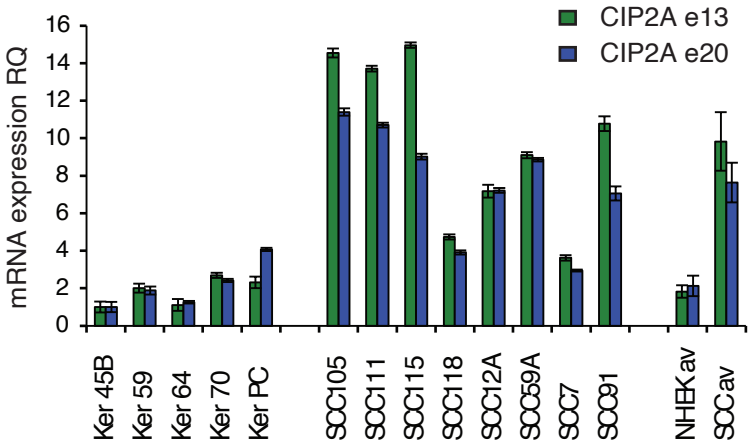


Supplemental Figure 3

A



B



Supplemental Figure 4

Table S1. Primer and probe sequences used in this study for PCR based analysis

RQ-PCR

Target	Forward primer	Probe	Reverse primer
NOCIVA#1	cagcctcactgaacatggaa	P34	cagtcaaaatggtgggaagg
NOCIVA#2	aaaagtggtctgtaaagctgctc	P49	ttcatgtcagattaccagtttatgc
CIP2A (e13)	cagtctggactgagaatattattgga	tccactgc*	ggcattgtttgctgctatacttt
CIP2A (e20)	gaacagataagaaaagagttgagcatt	cttcctcc*	cgaccttctaattgtgcctttt
b-actin	tcacccacacrgtgcccatctacgc	atgccctcccccatgccatcctgcgt	cagcggaaccgctcattgccaatgg
GAPDH	acccactcctccacctttga	acgaccactttgtcaagctcatttcctggt	ttgctgtagccaaattcgttgt

* Roche Universal ProbeLibrary (UPL) probe

For primer information of RQ-PCR of ARPP19, TIPRL, PME1, SET, EVI1, WT1 conducted on AML samples, please see Mäkelä et al, 2019, Cancers (Basel): Arpp19 Promotes Myc and Cip2a Expression and Associates with Patient Relapse in Acute Myeloid Leukemia

RQ-PCR for SPIRIT 2 CML samples

Target	Forward primer	Probe	Reverse primer
NOCIVA#3	ATGCCAAGACACAGTCAAAATG	CAGAGGCAGAGGATAA**	CCTGCTTGCATAAACTGGTAATC
GAPDH	ACCCACTCCTCCACCTTTGA	ACGACCACTTTGTCAAGCTCATTTCCTGGT**	TTGCTGTAGCCAAATTCGTTGT

**MGB Probe 6-FAM- seq -MGB-Eclipse®3`

RACE, PCR

Target	Primer sequence
CIP2A exon13 (GSP1)	tacttcaggacccacgtttgattact
CIP2A exon12 (GSP2)	cattggttcctggtatggaagtaagc
CIP2A exon6 (GSP3)	CGATAAAAAGATTTCACA
CIP2A exon7-8 (GSP4)	GCTCATATCTGGTG
CIP2A exon13 F2	CTGCTCCACTGCCAGATTT
CIP2A exon13 F5	TCAGGACCCACGTTTGATTAC
R1_NOCIVA	GCAGAGGATAAGACTTCCATGTT
R2_NOCIVA	ATAGAGCAGCTTTACAGACCAC
R3_NOCIVA	tgtaacttgaaaagcttcctgtgta
F1_CIP2A exon1	CCTGAATTCCCATGGACTCCACTGCCT
F2_CIP2A exon1	ATGGACTCCACTGCCT