International Clinical and Medical Case Reports Journal
Research Article (ISSN: 2832-5788)

Artificial Intelligence-Predicted Impact of Sequence Variants on Post-Translational Protein Modifications in MAP Kinase Kinases MAP2K1/2 in Rasopathies

Alex Sobko*

Alex Sobko, 8762728, Ofakim, Israel. Previously at The Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, 76100, Israel. Email: sobkosasha@gmail.com

Citation: *Alex Sobko. Artificial Intelligence-Predicted Impact of Sequence Variants on Post-Translational Protein Modifications in MAP Kinase Kinases MAP2K1/2 in Rasopathies. Int Clinc Med Case Rep Jour. 2023;2(2):1-10.*

Received Date: 13 January, 2023; **Accepted Date:** 18 January, 2023; **Published Date:** 20 January, 2023

***Corresponding author:** Alex Sobko. Alex Sobko, 8762728, Ofakim, Israel. Previously at The Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, 76100, Israel. Email: sobkosasha@gmail.com

ABSTRACT

We characterized dynamic localization, PTMs and activities of MAP2K1/2 homologue in social slime mould Dictyostelium (DdMEK1), in which MAP kinase pathways play central role in directional cell migration (chemotaxis) and development. DdMEK1 is subject to SUMOylation and Ubiquitination in response to chemoattractant stimulation. We also identified SUMO-directed RING finger Ubiquitin ligase, which forms a complex with SUMOylated DdMEK1 and ubiquitinates this protein kinase. Specific lysine residue in DdMEK1, modified by SUMO, was mapped, and SUMOylation-deficient mutant was characterized and shown to affect DdMEK1 subcellular localization, cell motility and multicellular development. More recently, MAP2K1 SUMOylation was demonstrated in mammalian cells and the mutation eliminating MAP2K1 SUMOylation was shown to be associated with cancer phenotype. These mutations can be associated with de novo and inherited neurodevelopmental syndromes called Rasopathies, that exhibit the defects in cell proliferation, growth, and invasiveness. We applied machine learning Artificial Intelligence (AI) modeling approach developed by the laboratory of Dr. Jüri Reimand PhD, OICR Informatics and Biocomputing, to characterize and validate disease-associated mutations that affect MAP2K1/2 PTMs (https://activedriverdb.org). Certain sequence variants in MAP2K1/2 genes

identified in the patient's biobank samples world-wide were predicted to affect MAP2K1/2 SUMOylation and/or Ubiquitination. In our ongoing and future work, we set to study patient-derived IPSC-based, and non-transformed cells engineered to possess particular mutations in order to examine how individual mutations in MAP2K1/2 genes affect kinase localization, activity and PTMs (SUMOylation, Ubiquitination). We plan to apply base/prime CRISPR gene editing strategies ("PTM modifiers") in order to eliminate those sequence variants that affect MAP2K1/2 SUMOylation and/or Ubiquitination and characterize cellular phenotypes.

Keywords: RASopathies; Ras-MAP kinase pathway; mutations; Single Nucleotide Polymorphisms (SNPs); sequence variants; SUMOylation; Ubiquitination; Post-translational protein modifications (PTMs)

INTRODUCTION

Mutations in the genes of Ras-MAP kinase pathway were discovered to play etiologic driver roles in many cancers and genetic neurodevelopmental disorders, called "RASopathies". Cardio-facio-cutaneous (CFC) syndrome and other conditions with mutations in the genes of canonical Ras-MAP kinase pathway (RASopathies) are sporadic genetic developmental disorders characterized by distinctive craniofacial features, including dental defects, heart abnormalities, mental retardation and ectodermal abnormalities. Germline hereditary and de-novo mutations in the genes of Ras-MAP kinase pathway (K-Ras, B-Raf, LZTR1, MAP2K1 and MAP2K2) were discovered to play etiologic roles in CFC and Noonan syndromes[1]. We would like to apply data mining and systems biology approaches in order to characterize these mutations and their effects on the activity of MAPK pathway.

In our previous molecular genetics, cell biology and biochemistry studies we worked on *Dictyostelium discoideum,* genetically tractable model organism, in which MAP kinase pathway is central in control of aggregation and directed cell migration. MAP kinase kinase (MAPKK homologue) is required for proper aggregation in *Dictyostelium.* Null mutations produce extremely small aggregate sizes, resulting in the formation of slugs and terminal fruiting bodies that are significantly smaller than those of wild-type cells. We elucidated a pathway regulating the localization and function of DdMEK1.

DdMEK1 is rapidly and transiently SUMOylated and ubiquitinated in response to chemoattractant, cAMP. SUMOylation is required for MEK1's function and its translocation from the nucleus to the cytosol and cortex, including the leading edge of chemotaxing cells. DdMEK1 in which the site of SUMOylation is mutated is retained in the nucleus and does not complement the mek1 null phenotype. Constitutively active DdMEK1 is cytosolic and is constitutively SUMOylated, whereas the corresponding non-activatable DdMEK1 is not SUMOylated and nuclear. MEK1 is also ubiquitinated in response to signaling. A MEK1-interacting, ubiquitin E3 ligase RING domain-containing protein controls nuclear localization and MEK1 ubiquitination. We studied the composition and function of this signaling complex that consists of DdMEK1 and associated DdMEK1-interacting protein (MIP1). We discovered first founding member of the family of SUMO-

directed RING Finger Ubiquitin Ligases (MIP1), in which RING domain serves as functional ubiquitin ligase module, which ubiquitinates MEK1 in vivo and in vitro[2,3,5].

Several years later, a study confirmed that MEK1 (MAP2K1) SUMOylation occurs also in mammalian cells, not only in *Dictyostelium*. This study implicated MEK1 SUMOylation in oncogenesis [4].

RESULTS

In the present study we set to use power of artificial intelligence approach to screen and search genomic databases and data biobanks for mutations and sequence variants, which affect SUMOylation and Ubiquitination of MAP2K1/2 proteins.

Laboratory of Dr. Jüri Reimand PhD, Informatics and Biocomputing Ontario Institute for Cancer Research (OICR) established publicly accessible database of the cancer-associated mutations that affect PTMs (https://activedriverdb.org). This database is based on data mining and machine learning. On March 23, 2021 they published major update containing new types of PTMs, including SUMOylation and Ubiquitination. We applied this machine learning Artificial Intelligence (AI) modeling approach to characterize and validate disease-associated mutations that affect MAP2K1/2 PTMs.

Certain sequence variants in MAP2K1/2 genes identified in the patient's biobank samples world-wide were predicted to affect MAP2K1/2 SUMOylation and/or Ubiquitination. These samples were found in patients with certain Rasopathies and tumor samples of the patients with several types of cancer. We hypothesize that sequence variants affecting PTMs such as SUMOylation and Ubiqutination in MAP2K1/2 play significant roles in etiology of these pathological conditions.

SUMOylation and Ubiquitination of MAP2K1/2.

According to the data of recent proteome-wide surveys of Ubiquitination sites (CST's "PhosphoSitePlus" database, www.phosphosite.org), 12 ubiquitin attachment lysine residues have been identified in MAP2K1 protein, and 14 sites – in MAP2K2 (Table 1). One of these sites (K104) is also known as an attachment site for SUMO1. These data are in good agreement with our previous hypothesis that SUMOylation sites functions as molecular switch coordinating protein localization and activity.

Sequence variants predicted to affect PTMs in MAP2K1/2.

The mutations in ClinVar, TCGA and PCAWG databases catalogued by "ActiveDriver" that, according to predictions of AI, might affect activities of PTM sites in MAP2K1 and MAP2K2 can be formally classified as those that target PTMs either ***directly*** or ***indirectly*** (through predicted intra- and inter-molecular

interactions and possible effects on the protein conformation). To date the only class of mutations ***directly*** affecting PTM in MAP2K1 is one that targets Ubiquitin attachment site K57 (Supplementary (Tables 1-3). No ***direct*** causative mutations associated with established disease phenotypes are known to eliminate or enhance SUMOylation of MAP2K1/2. We cannot exclude that the future data of ongoing patients' genome sequencing projects (exome, RNAseq and Whole Genome Sequencing) will identify such mutations in bio-samples of the patients with CFC and Noonan syndromes or certain types of cancers, but they are presently unknown [1].

We focused our attention on such patient-specific sequence variants that are predicted to affect SUMOylation and Ubiquitination of MAP2K1/2 – ***indirectly***, that are located either proximally or distally to the given PTM site (Supplementary (Tables 1, 2, 3). Driver mutation potential of these sequence variants has not yet been investigated and what we currently know is based solely on AI prediction and power of "ActiveDriver" database.

For example, E102G, R108I, I111N in MAP2K1 are predicted to affect K104 SUMOylation/Ubiquitination site in (ClinVar database) samples of the patients with CFC syndrome and possibly other RASopathies (Supplementary Table 1). I103N in MAP2K1 of the patients with cervical cancer (CESC), (TCGA database) also is expected to target K104 SUMOylation and Ubiquitination. All other listed mutations associated with CFC syndrome, other RASopathies and several types of cancer target individual Ubiquitination sites ***indirectly*** (see the data in Supplementary Tables 1-3).

Table 1: Ubiquitination (Ub) and SUMOylation (SUMO-1) PTM sites identified in human MAP2K1 and MAP2K2 proteins (Source: "PhosphositePlus" database v 6. 6. 0. 4).

MAP2K1	**MAP2K2**	**References**
[Lysine (K) that serves as PTM attachment site]	[Lysine (K) that serves as PTM attachment site]	
	K 6Ub	See: „Data references" for details [13].
K 36Ub	K 40Ub	
K 57Ub	K 51Ub	
K 64Ub	K 61Ub	
K 70Ub	K 63Ub	
K 88Ub	K 88Ub	
K 97Ub	K 101Ub	
K 104 SUMO-1	**K 108 SUMO-1**	

K 104Ub	K 108Ub
K 168Ub	K 163Ub
K 175Ub	K 196Ub
K 175Ac	K 209Ub
K 192Ub	K 361Ub
K 205Ub	K 370Ub
K 353Ub	K 385Ub

Supplementary Information (based on the MAP2K1/2 entries in "ActiveDriver" database).

(1). Suppl. Table 1: Mutations in MAP2K1 that affect Ubiquitination and SUMOylation ("ClinVar" database).

Mutation	Pos	Ref	Alt	Count	PTM impact	PTMs affected	Kinases	Disease annotations
MAP2K1 L42F	42	L	F	1	distal	1, K36Ub		Cardio-facio-cutaneous syndrome
MAP2K1 F53S	53	F	S	2	distal	1, K57Ub		Cardiofaciocutaneous syndrome 3, Cardio-facio-cutaneous syndrome
MAP2K1 K57Q	57	K	Q	2	direct	2, K57Ub; K64Ub		Cardio-facio-cutaneous syndrome, Rasopathy
MAP2K1 K57N	57	K	N	1	direct	2, K57Ub; K64Ub		Cardio-facio-cutaneous syndrome
MAP2K1 D67N	67	D	N	4	distal	2, K64Ub; K70Ub		Noonan syndrome, Cardiofaciocutaneous syndrome 3, Cardio-facio-cutaneous syndrome, Rasopathy
MAP2K1 L92R	92	L	R	1	distal	2, K88Ub; K97Ub		Cardio-facio-cutaneous syndrome
MAP2K1 E102G	**102**	**E**	**G**	**1**	**proximal**	**2, K97Ub; K104Ub; K104Sm**		**Cardiofaciocutaneous syndrome 3**
MAP2K1 R108L	**108**	**R**	**L**	**2**	**distal**	**1, K104Ub; K104Sm**		**Cardiofaciocutaneous syndrome 3, Rasopathy**
MAP2K1 I111N	**111**	**I**	**N**	**1**	**distal**	**1, K104Ub; K104Sm**		**Rasopathy**
MAP2K1 E203G	203	E	G	2	proximal	1, K205Ub		Noonan syndrome, Rasopathy

(2). Suppl. Table 2: Mutations in MAP2K1 that affect Ubiquitination and SUMOylation ("Cancer TCGA" database).

Mutation	Pos	Ref	Alt	Count	PTM impact	PTMs affected	Kinases	Cancer types
MAP2K1 F53I	53	F	I	1	distal	1, K57Ub		SKCM
MAP2K1 F53L	53	F	L	2	distal	1, K57Ub		LUAD, THYM
MAP2K1 F53V	53	F	V	1	distal	1, K57Ub		READ
MAP2K1 Q56P	56	Q	P	2	proximal	1, K57Ub		SKCM, COAD
MAP2K1 K57T	57	K	T	1	direct	2, K57Ub; K64Ub		LUAD
MAP2K1 K57N	57	K	N	4	direct	2, K57Ub; K64Ub		LUAD, HNSC, SKCM
MAP2K1 D67Y	67	D	Y	1	distal	2, K64Ub; K70Ub		ESCA
MAP2K1 S86A	86	S	A	1	proximal	1, K88Ub		BRCA
MAP2K1 R96K	96	R	K	1	proximal	1, K97Ub		LUSC
MAP2K1 I103N	**103**	**I**	**N**	**1**	**proximal**	**2, K97 Ub; K104 (Ub; Sm)**		**CESC**
MAP2K1 L177M	177	L	M	1	proximal	1, K175Ub; Ac		STAD
MAP2K1 E203K	203	E	K	4	proximal	1, K205Ub		SKCM, STAD, BRCA
MAP2K1 E203V	203	E	V	1	proximal	1, K205Ub		SKCM
MAP2K1 R349K	349	R	K	1	distal	2, K344Ub; K353Ub		CESC
MAP2K1 D351E	351	D	E	1	proximal	2, K344Ub; K353Ub		LUSC
MAP2K1	**103**	**I**	**N**	**1**	**proximal**	**1, K104Sm**		**CESC**

I103N								

(3). Suppl. Table 3: Mutations in MAP2K1 that affect Ubiquitination ("Cancer-PCAWG" database).

Mutation	Pos	Ref	Alt	Count	PTM impact	PTMs affected	Kinases	Cancer types
MAP2K1 F53L	53	F	L	2	distal	1, K57Ub		Panc-AdenoCA, Lymph-BNHL
MAP2K1 F53C	53	F	C	1	distal	1, K57Ub		Lymph-CLL
MAP2K1 K57T	57	K	T	1	direct	2, K57Ub; K64Ub		Lymph-BNHL
MAP2K1 K57N	57	K	N	1	direct	2, K57Ub; K64Ub		Lung-AdenoCA

DISCUSSION

According to our hypothesis, MAP2K1/2 is SUMOylated and then recruits SUMO-targeted Ubiquitin ligase (such as MIP1 in *Dictyostelium*), that then ubiquitinates this protein. We extensively studied in vivo dynamics of these PTMs in *Dictyostelium* in response to chemoattractant stimulation and found that these PTMs are transiently induced by this stimulus. Nothing is presently known about the stimuli regulating SUMOylation and Ubiquitination of human MAP2K1/2. Therefore, no data is available which of the identified PTMs are constitutive or inducible and what is the kinetics of these modifications. We also do not know whether K48-, K63- or other mono- or branched poly-Ubiquitin chains are involved in these PTMs and whether and how exactly they might affect kinase localization and activity. Given that most of the described PTM sites are located either within or close to the kinase domain, it is reasonable to predict that at least some of them might have an impact on kinase activities of MAP2K1/2 enzymes.

Based on the findings of Kubota et al.,[4] MAP2K1/2 SUMOylation is expected to suppress the activity of Ras-MAP kinase pathway, while the lysine-to-arginine mutation (K104R) might have a potential of oncogenic driver. Suppressive effect of MAP2K1/2 SUMOylation is in line with our data on recruitment of SUMO-targeted Ubiquitin ligase (such as RNF4 or its homologue) to SUMOylated MAP2K1/2, that leads to subsequent ubiquitination of the protein (possibly on the multiple lysine residues). If the scenario proposed in our *Dictyostelium* model is correct, SUMO-targeted Ubiquitin ligase functions by trapping modified MAP2K proteins (likely, within the nuclear compartment). K104R mutation, in contrast, might mimic de-SUMOylation that is expected to enhance the activity of Ras-MAP kinase pathway, and have oncogenic and Rasopathies-triggering potential. Future studies are expected to investigate the stimuli, kinetics and biological functions of individual PTM sites in MAP2K1/2.

CONCLUSION AND FUTURE DIRECTIONS

In our ongoing and future work, we set to apply "base-editing" and "prime-editing" strategies (CRISPR-Cas9 gene editing), [10, 11] to design and apply new generation of "PTM modifiers", that consist of Cas fused to additional molecule that results in precise modifications in the target PTM sites (Figure 1) Borrowed from [12], [12]. We plan to apply such "PTM modifiers" to patient-derived IPSC-based cell cultures and organoids, and also to non-transformed cells in order to substitute lysine residues, which are sites of SUMOylation and Ubiquitination, with arginine residues an modify PTM-affecting sequence variants in MAP2K 1/2 genes. To do so, we plan to use A-G base editor [7, 8, 9, 10,11] to convert AAA codon (K104) to AGA codon (R104) in the MAP2K1 gene. In the same fashion, by single nucleotide substitutions, we'll make specific sequence variants-bearing sequences. To ease transgene detection, prime-editor [10,11] will be engineered to fuse either fluorescent or small peptide tag in frame with the protein-encoding sequence in either N- or C-termini of the protein sequence. Using this combined base-/prime-editing strategy, we can either make detectable sequence variants associated with the known disease phenotypes (listed in "ActiveDriver" database), or vice versa, "cure" patient-derived IPSCs with such disease-associated sequence variants by substituting single nucleotides to revert them to the wild-type sequences.

The goal of this molecular engineering "tour de force" will be to characterize phenotypes of SUMOylation- and Ubiquitination-deficient mutants – examine MAP2K1/2 dynamic localization, cell viability, proliferation and motility. One may predict that the cells, expressing SUMOylation- and Ubiquitination-deficient MAP2K1/2 would have detectable defects in growth, proliferation and motility. These studies are aimed at understanding the functions of individual PTMs within MAP2K1/2 and their contributions to disease etiology and clinical picture, personally for every individual sequence variants-bearing patient.

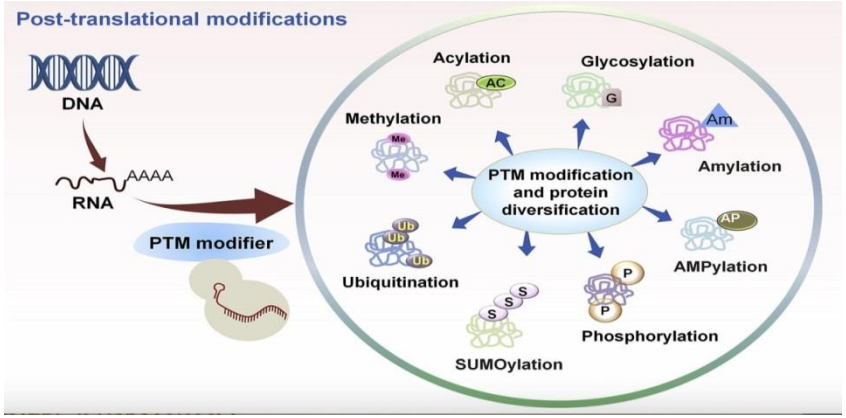

Figure 1: Use of Gene Editing (CRISPR Base-Editing; Prime-Editing) Tools in the Manipulation of Posttranslational Processes. Gene editing posttranslational modification (PTM) modifier consists of Cas fused to any additional molecule that results in precise modifications in the target PTM site. Borrowed from:[12] under: Creative Commons Attribution (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0/).

REFERENCES

1. Tartaglia M, Aoki Y, Gelb BD. The molecular genetics of RASopathies: An update on novel disease genes and new disorders. Am J Med Genet C Semin Med Genet. 2022.
2. Sobko A, Ma H, Firtel RA. Regulated SUMOylation and ubiquitination of DdMEK1 is required for proper chemotaxis. Dev Cell. 2002;2(6):745-756
3. Bussell K, On the move. Nat Rev Mol Cell Biol. 2002;(3):470.
4. Kubota Y, O'Grady P, Saito H, Takekawa M. Oncogenic Ras abrogates MEK SUMOylation that suppresses the ERK pathway and cell transformation. Nat Cell Biol. 2011;13(3):282-291.
5. Sobko A. A hypothetical MEK1-MIP1-SMEK multiprotein signaling complex may function in Dictyostelium and mammalian cells. Int J Dev Biol. 2020;64(10-11-12):495-498.
6. Krassowski M, Pellegrina D, Mee MW, Fradet-Turcotte A, Bhat M, Reimand J. ActiveDriverDB: Interpreting Genetic Variation in Human and Cancer Genomes Using Post-translational Modification Sites and Signaling Networks (2021 Update). Front Cell Dev Biol. 2021;9:626821.

7. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533. 2016;420–424.
8. Gaudelli NM. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature 551. 2017;464–471.
9. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-157.
10. Chen PJ, Liu DR. Prime editing for precise and highly versatile genome manipulation. Nat Rev Genet. 2022.
11. Sobko A. Principles and Significance of Base Editing and Prime Editing. J Mol Genet Med 16. 2022;572.
12. Pramanik D, Shelake RM, Kim MJ, Kim JY. CRISPR-Mediated Engineering across the Central Dogma in Plant Biology for Basic Research and Crop Improvement. Mol Plant. 2021;14(1):127-150.
13. Akimov V, Barrio-Hernandez I, Hansen SVF, Hallenborg P, Pedersen AK, Bekker-Jensen DB, et al. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. Nat Struct Mol Biol. 2018;25(7):631-640.
14. Udeshi ND, Svinkina T, Mertins P, Kuhn E, Mani DR, Qiao JW, et al. Refined preparation and use of anti-diglycine remnant (K-ε-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. Mol Cell Proteomics. 201;12(3):825-831.
15. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell. 2011;44(2):325-340.
16. Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, et al. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods. 2013;10(7):634-637.
17. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43:D512-520.

DATA REFERENCES

1. https://www.phosphosite.org/proteinAction.action?id=619&showAllSites=true
2. https://www.phosphosite.org/proteinAction.action?id=719&showAllSites=true
3. https://www.activedriverdb.org/sequence/show/NM_002755?filters=Site.types:in:ubiquitination
4. https://www.activedriverdb.org/sequence/show/NM_002755?filters=Mutation.sources:in:PCAWG;Site.types:in:ubiquitination
5. https://www.activedriverdb.org/sequence/show/NM_002755?filters=Mutation.sources:in:ClinVar;Site.types:in:ubiquitination;Mutation.gold_stars:in:4,3,2,1,1,0