

UniprotKB ID	Entry name	organism	full name	oglcnacscore	oglcnac sites	phosphorylation sites	PMIDS	sequence
O88491	NSD1_MOUSE	Mus musculus	Histone-lysine N-methyltransferase, H3 lysine-36 specific	15.818779	T2276;S2406	S110;S118;S380;S383;S1408;S2267;T2360;S2369	36852467	MDRTCELSRRNCLLSFSNPVNLDAS EDKDSPFGNGQSNFSEPLNGCTMQ LPTAASGTSQNAYGQDSPSCYIPLRR LQDLASMINVEYLSGSADGSESFQD PAKSDSRAQSPIVCTSLSPGGPTALA MKQEPTCNSPELQLRVTKTTKNGF LHFENFTGVDDADVDEMDPEQPV TEDESIEEIFEETQTNATCNYEPKSE NGVEVAMGSEQDSMPESRHGAVER PFLPLAPQTEKQKNKQRSEVDGSNE KTALLPAPTSLGDTNVTVEEQFNSIN LSFQDDPDSSPSPLGNMLEIPGTSSP STSQELPFVPOKILSKWEASVGLAEQ YDVPKGSKNQKCVSSSVKLDSEEDM PFEDCTNDPDSEHLLLNGCLKSLAF DSEHSADEKEKPCAКСRVRKSSDNI KRTSVKKDLVPFESRKEERRGKIPD NLGLDFISGGVSDKQASNELSRIANS LTGSSTAPGSFLFSSSVQNTAKTDFE TPDCDSL SGLSESALISKHSGEKKKL HPGQVCSSKVQLCYVGAGDEEKRS NSVSVSTTSDDGCSLDLDPTEHNSGF QNSVLGITDAFDKTENALSVHKNET QYSRYPVTNRIKEKQKSLITNSHADH LMGSTKTMEPETAELSQVNLSDLKI SSPIPKPQPEFRNDGLTTKFSAPPGI RNENPLTKGGLANQTLPLKCRQPK FRSIKCKHKESPAVAETSATSEDL KCCSSDTNGSPLANISKSGKGEGLK LLNNMHEKTRDSSDIETAVVKHVLS ELKELSYRSLSEDVSDSGTAKASKPL LFSSASSQNHIEPDYKFSTLLMML KDMHDSKTKEQRLMTAQNLASYRT PDRGDCSSGSPVGT SKVLVLGSSTP NSEKPGDSTQDSVHQSPGGGDSALS GELSSSLASDKREL PACGKIRSN CIPRRNCGRAPSSKLRETISAQMVK PSVNPKALKTERKRKFSRLPAVTLAA NRLGNKESGSVNGPSRGAEDPGK EPLQQMDLLRNEDTHFSDVHFDS KAKQSDPDKNLEKEPSFENRKGPEL GSEMNTENDELHGVNQVVPKRW QRLNQRRPKPGKRANRFREKENSE GAFGVLLPADAVQKAREDYLEQRAP PTSKPEDSAADPNHGSHSESVAPRL NVCEKSSVGMGDVEKETGIPSLMPQ TKLPEPAIRSEKKRLRKPSKWLEYT EEYDQIFAPKKKQKKVQEQVHKVSS RCEDESLLARCQPSAQNKQVDENSL ISTKEEPPVLEREAPFLEGPLAQSDL

GVTHAELPQLTSLVPAVEASPRPAL
ESEELLVKTPGNYESKRQRKPTKLL
ESNDLDPGFMPKKGDLGLSRKCFEA
SRSGNGIVESRATSHLKEFSGGTTKI
FDKPRKRKRQRLVTARVHYKKVKKE
DLTKDTPSSEGELLIHRTAASPKEILE
EGVEHDPGMSASKKLQVERGGGAA
LKENVCQNCEKLGELLCEAQCCGA
FHLECLGLPEMPRGKFCNECHTGI
HTCFVCKQSGEDVKRCLLPLCGKFY
HEECVQKYPPTVTQNKGFRCPLHICI
TCHAANPANVSASKGRLMRCVRCP
VAYHANDFCLAAGSKILASNSIICPN
HFTPRRGCRNHEHVNVSWCFVCSE
GGSLCCDSCPAAFHRECLNIDIPEG
NWYCNDCKAGKKPHYREIVWVKVG
RYRWWPAEICHPRAVPSNIDKMRH
DVGEFPVLFFGSNDYLWTHQARVFP
YMEGDVSSKDKMGKVDGTYKKAL
QEAARFEELKARKELRQLQEDRKN
DKKPPPYKHIKVNRPGRVQIFTADL
SEIPRCNCKATDENPCGIDSECINR
MLLYECHPTVCPAGVRCQNQCFSK
RQYPDVEIFRTLQRGWGLRKTDIKK
GEFVNEYVDELIDEECRARIRYAE
HDITNFYMLTLDKDRIIDAGPKGNYA
RFMNHCCQPNCEQKWSVNGDTR
VGLFALSDIKAGTELTFFNYNLECLGN
GKTVCKCGAPNCSGFLGVRPKNQPI
VTEEKSRKFKRPHGKRRSQGEVTK
EREDECFCGDAGQLVSCCKPGCPK
VYHADCLNLTKRPAGKWECPWHQC
DVCGKEAASFCEMCPSSFCKQHRE
GMLFISKLDGRLSCTEHDPCGPNPL
EPGEIREYVPPTATSPSPGTQPKEQ
SSEMATQGPKKSDQPPTDATQLLPL
SKKALTGSCQRPLLPERPPERTDSSS
HLLDRIRDLAGSGTKSOSLVSSQRP
QDRPPAKEGPRPQPPDRASPMTRPS
SSPSVSSLPLERPLRMTDSRLDKSIG
AASPQSQAVEKTPASTGLRLSSPDRL
LTTNSPKQISDRPPEKSHASLTQRL
PPPEKVL SAVVQSLVAKEKALRPVDQ
NTQSKHRPAVMDLIDLTPRQKERA
ASPQEVTPQADEKTAMLESSSWPSS
KGLGHIPRATEKISVSESLQPSGKVA
APSEHPWQAVKSLTHARFLSPPSAK
AFLYESATQASGRTPVGAEQTPGPPS
PAPGLVKQVKQLSRGLTAKSGQSFR
SLGKISASLPNEEKLTTEQSPWGL
GKASPGAGLWPIVAGQTLAQACWSA
GGTQTLAQTCWSLGRGQDPKPENAI
QALNQAPSSRKCADSEKK

