**Final Assessment – Drug Discovery in real life**

A new pharmaceutical company, Holien-X, is interested in finding new therapeutics for a rare type of children’s cancer, Neuroblastoma. They have gathered over 2000 patients with Neuroblastoma and have also managed to find genetically matched control patients and wish to sequence both to obtain the largest genetic set of data for this disease. Being a rare disease, these patients are very difficult to find, thus the method needs to be as accurate as possible.

**Your first step is to decide which technique you will use to sequence this data? To convince Holien-X, you must list why you chose this method and any pros/cons of this method (up to ½ page and/or table/figures).**

The sequencing was successful and using differential expression analysis (i.e. comparison between the genes upregulated in Neuroblastoma patients compared to the control patients) Holien-X has discovered the following 3 upregulated targets:

>sp|Protein1

MPSCSTSTMPGMICKNPDLEFDSLQPCFYPDEDDFYFGGPDSTPPGEDIWKKFELLPTPP

LSPSRGFAEHSSEPPSWVTEMLLENELWGSPAEEDAFGLGGLGGLTPNPVILQDCMWSGF

SAREKLERAVSEKLQHGRGPPTAGSTAQSPGAGAASPAGRGHGGAAGAGRAGAALPAELA

HPAAECVDPAVVFPFPVNKREPAPVPAAPASAPAAGPAVASGAGIAAPAGAPGVAPPRPG

GRQTSGGDHKALSTSGEDTLSDSDDEDDEEEDEEEEIDVVTVEKRRSSSNTKAVTTFTIT

VRPKNAALGPGRAQSSELILKRCLPIHQQHNYAAPSPYVESEDAPPQKKIKSEASPRPLK

SVIPPKAKSLSPRNSDSEDSERRRNHNILERQRRNDLRSSFLTLRDHVPELVKNEKAAKV

VILKKATEYVHSLQAEEHQLLLEKEKLQARQQQLLKKIEHARTC

>sp|Protein2

MASGSCQGCEEDEETLKKLIVRLNNVQEGKQIETLVQILEDLLVFTYSERASKLFQGKNI

HVPLLIVLDSYMRVASVQQVGWSLLCKLIEVCPGTMQSLMGPQDVGNDWEVLGVHQLILK

MLTVHNASVNLSVIGLKTLDLLLTSGKITLLILDEESDIFMLIFDAMHSFPANDEVQKLG

CKALHVLFERVSEEQLTEFVENKDYMILLSALTNFKDEEEIVLHVLHCLHSLAIPCNNVE

VLMSGNVRCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNILVLNEVHEFVVKAVQ

QYPENAALQISALSCLALLTETIFLNQDLEEKNENQENDDEGEEDKLFWLEACYKALTWH

RKNKHVQEAACWALNNLLMYQNSLHEKIGDEDGHFPAHREVMLSMLMHSSSKEVFQASAN

ALSTLLEQNVNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFEGSNTSLDIM

AAVVPKILTVMKRHETSLPVQLEALRAILHFIVPGMPEESREDTEFHHKLNMVKKQCFKN

DIHKLVLAALNRFIGNPGIQKCGLKVISSIVHFPDALEMLSLEGAMDSVLHTLQMYPDDQ

EIQCLGLSLIGYLITKKNVFIGTGHLLAKILVSSLYRFKDVAEIQTKGFQTILAILKLSA

SFSKLLVHHSFDLVIFHQMSSNIMEQKDQQFLNLCCKCFAKVAMDDYLKNVMLERACDQN

NSIMVECLLLLGADANQAKEGSSLICQVCEKESSPKLVELLLNSGSREQDVRKALTISIG

KGDSQIISLLLRRLALDVANNSICLGGFCIGKVEPSWLGPLFPDKTSNLRKQTNIASTLA

RMVIRYQMKSAVEEGTASGSDGNFSEDVLSKFDEWTFIPDSSMDSVFAQSDDLDSEGSEG

SFLVKKKSNSISVGEFYRDAVLQRCSPNLQRHSNSLGPIFDHEDLLKRKRKILSSDDSLR

SSKLQSHMRHSDSISSLASEREYITSLDLSANELRDIDALSQKCCISVHLEHLEKLELHQ

NALTSFPQQLCETLKSLTHLDLHSNKFTSFPSYLLKMSCIANLDVSRNDIGPSVVLDPTV

KCPTLKQFNLSYNQLSFVPENLTDVVEKLEQLILEGNKISGICSPLRLKELKILNLSKNH

ISSLSENFLEACPKVESFSARMNFLAAMPFLPPSMTILKLSQNKFSCIPEAILNLPHLRS

LDMSSNDIQYLPGPAHWKSLNLRELLFSHNQISILDLSEKAYLWSRVEKLHLSHNKLKEI

PPEIGCLENLTSLDVSYNLELRSFPNEMGKLSKIWDLPLDELHLNFDFKHIGCKAKDIIR

FLQQRLKKAVPYNRMKLMIVGNTGSGKTTLLQQLMKTKKSDLGMQSATVGIDVKDWPIQI

RDKRKRDLVLNVWDFAGREEFYSTHPHFMTQRALYLAVYDLSKGQAEVDAMKPWLFNIKA

RASSSPVILVGTHLDVSDEKQRKACMSKITKELLNKRGFPAIRDYHFVNATEESDALAKL

RKTIINESLNFKIRDQLVVGQLIPDCYVELEKIILSERKNVPIEFPVIDRKRLLQLVREN

QLQLDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIMAQILTVKVEGCPK

HPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLSDHRPVI

ELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYL

NWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDI

CGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLI

LADLPRNIMLNNDELEFEQAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQE

LVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVAD

GLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTPGFRA

PEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEYG

CAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNVIVECMVATHH

NSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESWIVSGTQ

SGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTV

KLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRT

SQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLREVMVKE

NKESKHKMSYSGRVKTLCLQKNTALWIGTGGGHILLLDLSTRRLIRVIYNFCNSVRVMMT

AQLGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHIEVRKELAEK

MRRTSVE

>sp|Protein3

MNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALP

RVTPFYAVKCNDSKAIVKTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCKQVSQI

KYAANNGVQMMTFDSEVELMKVARAHPKAKLVLRIATDDSKAVCRLSVKFGATLRTSRLL

LERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPG

SEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ

TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTC

DGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQF

QNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINV

**Identify and compare these 3 potential targets and decide which one you would choose for a drug discovery campaign. To convince Holien-X, you must list the pros/cons of this target and show evidence for why this target is the best for a drug discovery campaign (approx. ½ - 1 page + figures/ tables).**

Holien-X has re-analysed their data using network-based methods and discovered new information which shows that Aurora Kinase B (Uniprot code: Q96GD4) is the best druggable option as it is:

* highly expressed and upregulated in the diseased patients compared to controls
* has close homology to mouse
* has a close homolog to construct a high quality homology model
* has a known role in cancer

**Your job is to download the alphafold homology model and confirm its suitability for a drug discovery campaign. For example, you may like to; analyse the quality of this model (i.e. https://swissmodel.expasy.org/assess), analyse the properties of the protein, search for druggable pockets etc. Write these details into your report with figures to guide the team at Holien-X (approx. ½ - 1 page + figures/tables).**

Holien-X has taken your advice on board and would like to conduct a Virtual Screen. **Write a short proposal for the steps you will undertake in order to do this (up to ½ page and/or table/figures)**

Congratulation, your virtual screen was successful. Holien-X has screened the compounds you suggested and has identified 5 compounds which bind to the wild-type protein but not an A105R mutant isoform (confirming your active site). They also reduce the growth of Neuroblastoma cell culture.

**Analyse the following table and let Holien-X know which compound you would choose to develop further and why? (up to ½ page and/or table/figures)**

|  |  |
| --- | --- |
| **SMILES String** | **Activity (IC50)** |
| O=C(C1=CC=CC(Cl)=C1F)N(CC2)CCN2CC3=CC=CC(CC4=NC=CS4)=N3 | 1nM |
| CN(C)CC1=CC(C2=CC(C(C3=CN(CC)N=C3C4=CC=CC=C4)=NC=N5)=C5N2)=CC=C1 | 0.5nM |
| CCN1N=C(C2=CC=C(F)C=C2)C([C@H]3CCCC(C(F)(F)F)C3)=C1 | 1.5nM |
| O=C(NC1=CCC=C([C@H]2CCOC2)C1)C3=NC=CN=C3 | 1mM |
| O=[N+]([O-])C1=CC=C(C2=NNC(SCC#CC)=N2)C=C1C | 1mM |

Holien-X has now developed your compound into Phase 2 clinical trials. Unfortunately, they are finding a subset of patients which are showing resistance to the drug. They have sequenced these patients, and all have the following nucleotide sequence.

>MutantProtein

atggcgcagaaagaaaacagctatccgtggccgtatggccgccagaccgcgccgagcggc

ctgagcaccctgccgcagcgcgtgctgcgcaaagaaccggtgaccccgagcgcgctggtg

ctgatgagccgcagcaacgtgcagccgaccgcggcgccgggccagaaagtgatggaaaac

agcagcggcaccccggatattctgacccgccattttaccattgatgattttgaaattggc

cgcccgctgggcaaaggcaaatttggcaacgtgtatctggcgcgcgaaaaaaaaagccat

tttattgtggcgctgaaagtgctgtttaaaagccagattgaaaaagaaggcgtggaacat

cagctgcgccgcgaaattgaaattcaggcgcatctgcatcatccgaacattgaacgcctg

tataactatttttatgatcgccgccgcatttatctgattctggaatatgcgccgcgcggc

gaactgtataaagaactgcagaaaagctgcacctttgatgaacagcgcaccgcgaccatt

atggaagaactggcggatgcgctgatgtattgccatggcaaaaaagtgattcatcgcgat

attaaaccggaaaacctgctgctgggcctgaaaggcgaactgaaaattgcggattttggc

tggagcgtgcatgcgccgagcctgcgccgcaaaaccatgtgcggcaccctggattatctg

ccgccggaaatgattgaaggccgcatgcataacgaaaaagtggatctgtggtgcattggc

gtgctgtgctatgaactgctggtgggcaacccgccgtttgaaagcgcgagccataacgaa

acctatcgccgcattgtgaaagtggatctgaaatttccggcgagcgtgccgatgggcgcg

caggatctgattagcaaactgctgcgccataacccgagcgaacgcctgccgctggcgcag

gtgagcgcgcatccgtgggtgcgcgcgaacagccgccgcgtgctgccgccgagcgcgctg

cagagcgtggcg

**Holien-X would like to understand what the patient mutant is? Is it a modest mutation or significant? Where on the protein this mutation is occurring? Is it likely to influence compound binding or another aspect of the protein function? (½-1 page + table/figures**

Congratulations based on your analysis the drug has now passed all approvals and is being used to treat these patients. **Please add a summary sentence or two describing how you feel bioinformatics helped these patients.**