 doi:10.20944/preprints201711.0008.v1

ARTICLE TYPE

Revealing ETC-1922159 Affected Unknown 3^{rd} Order Combinations of DNA Repair & Genomic Stability Factor RAD51 Family, in Silico†

shriprakash sinha★,◕

DNA repair helps in maintaining the proper and healthy functioning for the cells in the human body. Failure in DNA repair process can lead to aberrations as well as tumorous stages. There are various types of damages that a DNA can go through, one of which is the DNA double strand breaks (DSB) that can be repaired via homologous recombination (HR). RAD51 plays a central role in HR and has been implicated as a negative/poor prognostic marker for colorectal adenocarcinoma, with high expression in colorectal cancer. Mechanistically, RAD51AP1 facilitates RAD51 during the repairing process by binding with RAD51 via two DNA binding sites, thus helping in the D-loop formation in the HR process. Often, in biology, we are faced with the problem of exploring relevant unknown biological hypotheses in the form of myriads of combination of factors that might be affecting the pathway under certain conditions. For example, RAD51AP1-XRCC2 is one such 2^{nd} order combination whose relation needs to be tested under the influence recently developed Porcupine-WNT inhibitor ETC-1922159. The x-ray repair cross complementing XRCC family is known to work as a mediator or stabilizer for RAD51 during the HR process. The inhibitor is known to suppress Porcupine and thus inhibit a range of oncogenes known to be directly or indirectly affected by the Wnts. In a recent unpublished work in bioRxiv, we had the opportunity to rank these unknown biological hypotheses for down regulated genes at 2^{nd} order level after the drug was administered. The in silico observations showed that the combination of RAD51AP1-XRCC2 was assigned a relatively lower rank, thus validating the pipeline's efficacy with the confirmed wet lab experiment that indicate that both RAD51AP1 and XRCC2 were down regulated after treatment in cancer cells. Here, we take one step further by in silico analysis of the 3^{rd} order combinations of RAD51-X-X & RAD51AP1-X-X (X can be known or unknown factor), from a range of 100 randomly picked down regulated genes after ETC-1922159 treatment. The pipeline uses the density based HSIC (Hibert Schmidth Information Criterion) sensitivity index with an rbf (radial basis function) kernel, which is known to be highly effective in sensitivity analysis. Various unknown/unexplored/untested RAD51/RAD51AP1 related 3^{rd} order biological hypotheses emerge some of which are confirmed in wet lab, while others need to be tested.
Keywords: WNT pathway; porcupine inhibitor ETC-1922159; sensitivity analysis; colorectal cancer; unknown biological hypotheses; combinatorial search space; support vector ranking; DNA repair and genomic stability factor RAD51

† *Aspects of unpublished work presented as poster in the recently concluded Wnt Gordon Conference, from 6-11 August 2017, held in Stowe, VT 05672, USA.*
★ Corresponding author contact: sinha.shriprakash@yandex.com
◕ *Worked as independent researcher*

Significance

DNA repair & genomic stability play a major role in maintaining the health of the cells. RAD51AP1 mechanistically facilitates

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2017 doi:10.20944/preprints201711.0008.v1

RAD51 during the repairing process by binding with RAD51 via two DNA binding sites, thus helping in the D-loop formation in the homologous recombination process. Failure in these can lead to aberrations and tumorous stages. The current work reveals at in silico level, the 3rd order RAD51AP1 associated combinations that might be affected by the administration of ETC-1922159. The potential of revealing such higher order unknown biological hypotheses via ranking is indispensable in the current era of search in a vast combinatorial search forest. It provides a guided navigation for the oncologists where cherry picking the usual norm for testing.

Introduction

DNA repair and genomic stability

DNA repair is an important aspect in maintaining the proper and healthy functioning for the cells in the human body. Failure in DNA repair process can lead to aberrations as well as tumorous stages. There are various types of damages that a DNA can go through, one of which is the DNA double strand breaks (DSB) that can be repaired via homologous recombination (HR). RAD51 plays a central role in HR and is known to function in the three phases of HR namely : presynapsis, synapsis and post-synapsis [?] . Recently, RAD51 has been implicated as a negative/poor prognostic marker for colorectal adenocarcinoma and has been found to be highly expressed[?] . A negative/poor prognostic marker indicates that it is harder to control the malignancy. Since RAD51 helps in the repair of the DNA damage via HR and is implicated as a poor prognostic marker in colorectal adenocarcinoma, this suggests its functionality in maintaining genomic stability and therapeutic resistance to cancer drugs[?] &[?] . Mechanistically, RAD51AP1 facilitates RAD51 during the repairing process by DNA binding via two DNA binding sites, thus helping in the D-loop formation in the HR process[?] &[?] .

RAD51

RAD51, an ortholog of Esherichia coli RECA[?] is found to be operating in mammals and work in three steps during the homologous recombination process. During the presynaptic phase there is a formation of the RAD51-ssDNA filament as RAD51 is loaded to form a stretch of the filment with single stranded DNA (ssDNA). This ssDNA is generated from degradation and deletion of unwound double strand of DSBs. During the synaptic phase there is a formation of the D-loop were RAD51 facilitates the intruding DNA to form a connection with the homologous duplex DNA. This leads to formation of RAD51-dsDNA (ds - double strand) from the intruding and the donar ssDNA. Finally, DNA synthesis takes place via DNA polymerases and an invading primer, during the postsynaptic phase when the RAD51 disassociates from the RAD51-dsDNA filament to leave the dsDNA homologous pair for further processing. A good survey of this process has been described in[?] . The formation of the RAD51 nucleoprotein filaments has been described in detail in[?] &[?] .

RAD51AP1

RAD51AP1 or RAD51 associated protein 1, enhances the activity of RAD51, by two DNA binding sites during the D-loop formation[?] . Mutant studies[?] show that both C-terminal and N-terminal binding sites are indispensable for RAD51AP1 to function along with RAD51. More specifically, RAD51AP1 associates with DMC1 mediated D-loop formation and[?] provide evidence that RAD51AP1 cooperates with the DMC1 presynaptic filament to capture duplex DNA and to assemble the synaptic complex, in which the recombining DNA molecules are joined paranemically. Thus, RAD51AP1 maintain genomic stability via RAD51 recombinase enhancement[?] .

PORCN-WNT inhibitors

The regulation of the Wnt pathway is dependent on the production and secretion of the WNT proteins. Thus, the inhibition of a causal factor like PORCN which contributes to the WNT secretion has been proposed to be a way to interfere with the Wnt cascade, which might result in the growth of tumor. Several groups have been engaged in such studies and known PORCN-WNT inhibitors that have been made available till now are IWP-L6[?] &[?] , C59[?] , LGK974[?] and ETC-1922159[?] . In this study, the focus of the attention is on the implications of the ETC-1922159, after the drug has been administered. The drug is a enantiomer with a nanomolar activity and excellent bioavailability as claimed in[?] .

Combinatorial search problem and a possible solution

We have already addressed the issue of combinatorial search problem and a possible solution in[?] and[?] . The details of the methodology of this manuscript have been explained in great detail in[?] & its application in[?] and readers are requested to go through the same for gaining deeper insight into the working of the pipeline and its use for published data set generated from ETC-1922159. In order to understand the significance of the solution proposed to the problem of combinatorial search that the biologists face in revealing unknown biological search problem, these works are of importance. Using the same code with minor modifications in[?] and[?] , it was possible to generate the rankings for 3^{rd} order combinations. 100 genes were randomly selected from the list of down regulated genes, by the pipeline and a 3^{rd} order combination was generated from those 100 genes. The total number of gene combination with $C_3^{100} = 161700$. Out of these the RAD51 and RAD51AP1 associated 3^{rd} order combinations were selected, which account to a total of 4851 combinations. The goal of this manuscript is to analyse some of these 3^{rd}

 doi:10.20944/preprints201711.0008.v1

order ranked RAD51 and RAD51AP1 associations.

Results and Discussion

We present here the 3rd order combinations associated with RAD51/RAD51AP1 and represent them as RAD51/RAD51AP1-X-X, were X can be known or unknown factor from a list of genes that were affected after the administration of the ETC-1922159 drug. There are a total of 4851 combinations of randomly selected 100 genes from a list of 2500± genes. Out of these 100, RAD51/RAD51AP1 was one of them. Here we analyse some of the ranked combinations out of 4851 3^{rd} order interactions for RAD51-X-X and RAD51AP1-X-X. Note that the rankings were generated using only the HSIC density index using the radial basis function kernel. Also, the rankings for a particular gene might change over different combinations and the biologists/oncologists are advised to cross check across the different tables presented. However, where possible, we report confirmatory results by the pipeline that fall in line with the published and known mechanism of a particular gene under consideration. Also, many of the combinations are yet to be tested and we make openings for the deeper analysis and exploration of the combinations as future work. These combinations with their rankings have been recored in table 1 and 2.

RAD51-X-X, X - known/unknown/untested factor

In table 1, we present the rankings of 3^{rd} order combination at particular intervals, for RAD51.

RAD51AP1-X-X, X - known/unknown/untested factor

In table 2, we present the rankings of 3^{rd} order combination at particular intervals, for RAD51AP1.

Conclusion

RAD51 and the complementing paralogue RAD51AP1, work in tandem to maintain genomic stability and play a greater role during homologous recombination process when double strand breaks are encountered. Here we present a range of ranking for both RAD51 and RAD51AP1 associated $3rd$ order combinations after the treatment of ETC-1922159 in colorectal cancer cells. Rankings guide oncologists/biologists to study the influence of these genomic stability and DNA repair factors in context of other factors in colorectal cancer case and the in silico pointers of prioritization facilitate in the emergence of some unknown/untested/unexplored biological hypotheses.

Conflict of interest

There are no conflicts to declare.

Acknowledgements

Special thanks to Mrs. Rita Sinha and Mr. Prabhat Sinha for supporting the author financially, without which this work could not have been made possible.

Source of Data

Data used in this research work was released in a publication in[?] . The ETC-1922159 was released in Singapore in July 2015 under the flagship of the Agency for Science, Technology and Research (A*STAR) and Duke-National University of Singapore Graduate Medical School (Duke-NUS).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2017 doi:10.20944/preprints201711.0008.v1

RANKING USING HSIC - RBF KERNEL W.R.T RAD51

3^{nd} odr comb.	rbf rank	3^{nd} odr comb.	rbf rank	3^{nd} odr comb.	
RAD51-DHODH-ART3	1	RAD51-XX232-ARV1	50	WDR46-RAD51-TRAPPC6A	100
RAD51-SKA1-ART3	150	RAD51-DHX30-NKD1	200	RAD51-PKD2-GCDH	250
RAD51-CD19-ART3	300	RAD51-MDH1-CENPE	350	RAD51-XX232-SHF	400
RAD51-MTIF2-SCML2	450	RAD51-TMEM243-XX1	500	RAD51-MTIF2-EXTL2	550
RAD51-MDH1-GCDH	600	RAD51-DHX30-DNAJC9.AS1	650	RAD51-MDH1-PYURF	700
RAD51-PCCB-TROAP	750	RAD51-CEP152-XX1	800	RAD51-CENPN-XX190	850
RAD51-PCCB-SCML2	900	RAD51-SLC19A1-XX1	950	RAD51-TEX2-XX1	1000
RAD51-CUEDC2-EXTL2	1050	RAD51-GCDH-DNAJC9.AS1	1100	RAD51-SNRPD3-SNRPF	1150
RAD51-CEP152-SKA1	1200	RAD51-PCCB-IFT122	1250	RAD51-SNRPD3-PYURF	1300
RAD51-TOMM20L-NXPE4	1350	RAD51-SCML2-SCARA3	1400	RAD51-TPM2-CKAP5	1450
RAD51-ALKBH8-SCARA3	1500	RAD51-XX49-TAF9	1550	RAD51-NKD1-XX190	1600
RAD51-CEP152-XX218	1650	RAD51-BIVM-WDR91	1700	RAD51-BCL7A-CACYBP	1750
RAD51-DNAJC9.AS1-ARV1	1800	RAD51-EIF4B-HSPD1	1850	RAD51-CCDC115-LRRC45	1900
RAD51-ERI2-SNHG16	1950	RAD51-MRPS25-PAXIP1.AS1	2000	RAD51-ERI2-XX53	2050
RAD51-SLC17A9-BOD1	2100	RAD51-PPIH-CUEDC2	2150	RAD51-C19orf54-SCARA3	2200
RAD51-UTP3-TEX2	2250	RAD51-IQCB1-XX53	2300	RAD51-PCCB-ACSM3	2350
RAD51-DIS3L2-TOPBP1	2400	RAD51-CDK5RAP1-XX232	2450	RAD51-SNRPD3-ERI2	2500
RAD51-CACYBP-BUB1B	2550	RAD51-HIBCH-C5orf34	2600	RAD51-NDNL2-SNHG16	2650
RAD51-TOMM20L-CEP57	2700	RAD51-ZCCHC7-EXTL2	2750	RAD51-HSPD1-ARSE	2800
RAD51-EXTL2-ARV1	2850	RAD51-IFT122-EIF4B	2900	RAD51-CENPN-CACYBP	2950
RAD51-UTP20-CENPE	3000	RAD51-XX68-CCDC115	3050	RAD51-GCDH-HIBCH	3100
RAD51-EXTL2-COQ10A	3150	RAD51-SLC17A9-COQ10A	3200	RAD51-ERI2-DIS3L2	3250
RAD51-PPIH-NKD1	3300	RAD51-XX218-TOPBP1	3350	RAD51-HNRNPA1-CKAP5	3400
RAD51-IFT122-MAGEF1	3450	RAD51-MRPL45-RHPN1	3500	RAD51-ARSE-RPS3A	3550
RAD51-PPIH-DNAJC9.AS1	3600	RAD51-CDK5RAP1-C19orf54	3650	RAD51-DNAJC9.AS1-ARSE	3700
RAD51-SMARCC1-TCF19	3750	RAD51-HNRNPA1-UTP20	3800	RAD51-ALG1L-RHPN1	3850
RAD51-NDNL2-DIS3L2	3900	RAD51-CDK5RAP1-DIS3L2	3950	RAD51-XX49-DIS3L2	4000
RAD51-EIF4B-XX53	4050	RAD51-SCML2-SNRPD3	4100	RAD51-NKD1-NDNL2	4150
RAD51-TPGS2-ERI2	4200	RAD51-RPS3A-TCF19	4250	RAD51-MRPS25-GPALPP1	4300
RAD51-SMARCC1-SCARA3	4350	RAD51-PPIH-TCF19	4400	RAD51-XX49-HIBCH	4450
RAD51-XX49-MKKS	4500	RAD51-ATAD5-MAGEF1	4550	RAD51-ARSE-DIS3L2	4600
RAD51-XX49-C19orf54	4650	RAD51-DNAJC9.AS1-MKKS	4700	RAD51-NXPE4-RPS3A	4750
RAD51-CDK5RAP1-TEX2	4800	RAD51-CENPE-PAXIP1.AS1	4850		

Table 1 3^{rd} order interaction ranking using HSIC for radial basis function kernel. Total number of 3^{rd} order interactions in a set of 100 genes - 161700. 4851 3^{rd} order combinations for RAD51 associated work. Rankings for RAD51-X-X have been tabulated.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2017 doi:10.20944/preprints201711.0008.v1

RANKING USING HSIC - RBF KERNEL W.R.T RAD51AP1

3^{nd} odr comb.	rbf rank	3^{nd} odr comb.	rbf rank	3^{nd}odr comb.	rbf rank
IMPACT-NMU-RAD51AP1	1	RPL6-NUBPL-RAD51AP1	50	C22orf39-TMEM109-RAD51AP1	100
IMPACT-POLE3-RAD51AP1	150	DHX35-XX107-RAD51AP1	200	POLE3-KIFC1-RAD51AP1	250
CCDC34-BIRC5-RAD51AP1	300	FBXL8-RAD51AP1-ZNRF3	350	ATPAF1-CENPO-RAD51AP1	400
ENOSF1-CCDC66-RAD51AP1	450	C22orf39-HSPB6-RAD51AP1	500	QDPR-RAD51AP1-ARHGAP11A	550
C22orf39-HOXB5-RAD51AP1	600	ZC3H8-EZH2-RAD51AP1	650	C19orf44-SS18L2-RAD51AP1	700
ZNF614-RAD51AP1-ARHGAP11A	750	HNMT-BIRC5-RAD51AP1	800	ATPAF1-BOD1-RAD51AP1	850
DHX35-QTRTD1-RAD51AP1	900	THOC1-NMU-RAD51AP1	950	XX98-CCDC34-RAD51AP1	1000
C22orf39-CCDC66-RAD51AP1	1050	NMU-RAD51AP1-NUDT1	1100	TIMMDC1-RAD51AP1-GOLGA2P5	1150
CCNI2-COL9A3-RAD51AP1	1200	TIMMDC1-CCNI2-RAD51AP1	1250	NUBPL-SLC12A8-RAD51AP1	1300
TBC1D16-RBL1-RAD51AP1	1350	RPL24-RAD51AP1-ARHGAP11A	1400	DHX35-BIRC5-RAD51AP1	1450
ERI3-FBXL8-RAD51AP1	1500	SS18L2-CCSAP-RAD51AP1	1550	TBC1D16-PRPS1-RAD51AP1	1600
POLE3-XX98-RAD51AP1	1650	HSPB6-KIFC1-RAD51AP1	1700	C19orf44-EZH2-RAD51AP1	1750
ACO1-RAD51AP1-ZNRF3	1800	SS18L2-ZC3H8-RAD51AP1	1850	ACO1-QDPR-RAD51AP1	1900
MAP10-CCDC66-RAD51AP1	1950	FANCG-MIS18A-RAD51AP1	2000	RPL7A-RPL13-RAD51AP1	2050
RRM1.AS1-PDSS1-RAD51AP1	2100	FBXL8-TCERG1-RAD51AP1	2150	RBL1-TRIM59-RAD51AP1	2200
HSPB6-EARS2-RAD51AP1	2250	MDN1-PRPS1-RAD51AP1	2300	FANCF-CENPE-RAD51AP1	2350
FANCG-ATPAF1-RAD51AP1	2400	DNMT1-CENPE-RAD51AP1	2450	TCEA3-DNMT1-RAD51AP1	2500
DDX12P-POLE3-RAD51AP1	2550	FKTN-METTL16-RAD51AP1	2600	DNMT1-ENOSF1-RAD51AP1	2650
TRIM59-CDK1-RAD51AP1	2700	SPN-CDK1-RAD51AP1	2750	MIS18A-HOXB5-RAD51AP1	2800
CDK1-PDSS1-RAD51AP1	2850	SOAT1-CCDC66-RAD51AP1	2900	RBL1-CDK5RAP2-RAD51AP1	2950
FAM221A-CCDC66-RAD51AP1	3000	SS18L2-FAM221A-RAD51AP1	3050	METTL16-CLDN2-RAD51AP1	3100
FANCG-RAD51AP1-ASF1B	3150	TCEA3-RPL4-RAD51AP1	3200	RMND1-RAD51AP1-OSGEPL1	3250
MIS18A-C4orf46-RAD51AP1	3300	TRIM59-CDK5RAP2-RAD51AP1	3350	QTRTD1-RAD51AP1-XX6	3400
SPN-CLDN2-RAD51AP1	3450	NUBPL-EZH2-RAD51AP1	3500	RAD51AP1-LCTL-XX6	3550
PRPS1-EARS2-RAD51AP1	3600	ZC3H8-ERI3-RAD51AP1	3650	HSPB6-RAD51AP1-NUDT1	3700
BRIP1-CCSAP-RAD51AP1	3750	XX181-CDK5RAP2-RAD51AP1	3800	CDK1-PRPS1-RAD51AP1	3850
FANCG-ACO1-RAD51AP1	3900	HIBCH-QDPR-RAD51AP1	3950	XX160-ATPAF1-RAD51AP1	4000
DCAF13-RAD51AP1-XX210	4050	DCAF13-BRIP1-RAD51AP1	4100	CCSAP-RAD51AP1-PAICS	4150
TCEA3-RAD51AP1-OSGEPL1	4200	MTR-NACA-RAD51AP1	4250	CGREF1-RAD51AP1-NUDT1	4300
RRM1.AS1-HSPB6-RAD51AP1	4350	EARS2-RAD51AP1-XX210	4400	SLC12A8-TPPP3-RAD51AP1	4450
DDX12P-TPPP3-RAD51AP1	4500	C4orf46-DDX12P-RAD51AP1	4550	TBC1D16-RAD51AP1-STIL	4600
IFT46-RAD51AP1-LCTL	4650	HSPB6-RAD51AP1-GRIN2B	4700	C4orf46-HIBCH-RAD51AP1	4750
EZH2-RAD51AP1-STIL	4800	SPN-RAD51AP1-STIL	4850		

Table 2 3^{rd} order interaction ranking using HSIC for radial basis function kernel. Total number of 3^{rd} order interactions in a set of 100 genes - 161700. 4851 3^{rd} order combinations for RAD51AP1 associated work. Rankings for RAD51AP1-X-X have been tabulated.