RESEARCH ARTICLE Indian Journal of Animal Research, Volume 54 Issue 9: 1086-1097 (September 2020)

Exploration of RNA-Guided Recombinase Target Sites for Hyperactivated Recombinase Beta in Bovine Genome

Shalu Kumari Pathak, Arvind Sonwane, Subodh Kumar

10.18805/ijar.B-3880

ABSTRACT

RNA-guided recombinases (RGR) are potentially valuable tools for basic research and genetic modifications. The platform has been demonstrated to do genome editing efficiently. The platform operates on a typical recognition site comprised of the degenerate recombinase site, a 5 to 6-base pair spacer flanking it and this whole central region is flanked by two guide RNA-specified DNA sequences or Cas9 binding sites which is followed by protospacer adjacent motifs. In present investigation, a detailed map of target sites for RNA-guided recombinase platforms based on hyperactivated recombinase Beta throughout the bovine genome was prepared. For this, Chromosome wise whole genomic sequence data was retrieved from Ensembl followed by designing search pattern for recombinase Beta with spacer length five. By using this search pattern, RGR target sites were located by using dreg program of Emboss package. In total,436 RGR target sites were identified in bovine genome for recombinase Beta with spacer length five. These RGR target site provide potential of being utilized for specific genomic integration, deletion or inversion.

Key words: Bovine genome, Dreg, Emboss, Hyperactivated recombinase Beta, RNA guided recombinase target sites.

INTRODUCTION

The advent of recent genome engineering technologies greatly enhanced the capabilities for investigating the genetic basis of disease and exploring the development of novel gene therapies. Recombinases are DNA manipulating enzymes which serve as powerful tools for predictable and specific genetic manipulations.Site-specific recombinases directly catalyze the cleavage, strand exchange and re-ligation of two double-stranded DNA sequences which results in the insertion, deletion or inversion of sequences of interest, depending on the relative orientation of the substrate sequences.Direct catalysis by recombinases typically does not provoke error-prone DNA repair processes that result in indel formation and also not dependent on endogenous cellular DNA repair machinery.Site-specific recombinases are highly specialized enzymes that promote DNA rearrangements between specific target sites.

Site-specific recombinases (SSRs) can be categorized within one of two structurally and mechanistically distinct groups: Tyrosine recombinases - e.g., Cre, Flp and the λ integrase (Grainge and Jayaram, 1999) and Serine reombinases - φC31 integrase, γδ resolvase and Gin invertase (Smith and Thorpe, 2002).The tyrosine recombinases break and rejoin pairs of single DNA strands to generate Holliday junction intermediates, while the serine recombinases cleave all four DNA strands before promoting strand exchange and re-ligation. Due to complex protein-protein and protein-DNA interactions required to coordinate catalysis, Tyrosine recombinases exhibit remarkable target site specificity. But, altering the specificity of many Tyrosine SSRs has proven difficult (Grindley *et al*. 2006; Sarkar *et al*. 2007). In this context, Serine recombinases of the resolvase/invertase type provide a versatile alternative to tyrosine recombinases for genome engineering. Native DBDs of serine recombinases can be replaced with custom-designed proteins to generate chimeric recombinases (Gaj *et al*. 2011; Gordley *et al*., 2009; Akopian *et al*. 2003).

SSRs are hampered by the strict recognition specificities for their natural DNA targets (Grindley *et al*., 2006). So, application of these enzymes has been limited to cells or organisms that contain rare pre-existing pseudo-recognition sites or pre-introduced target sites through time-consuming and labor-intensive procedures (Thyagarajan *et al*. 2000; Chalberg *et al*. 2006). Unique topological and spatial constraints are imposed onto these enzymes through the presence of multiple binding sites or accessory factor proteins that ensure the specificity of the recombination reaction (Sancar *et al*. 2004; Smith and Thorpe, 2002). In order for this technology to reach its full potential, Hyperactivated variants of recombinases have been developed by methods like directed evolution and rational design of custom recombinases.

To increase the number of sites amenable for targeted recombination in cells, researchers have fused hyperactive variants of small serine recombinases to zinc finger and TALE DNA-binding proteins (Akopian *et al*. 2003; Gordley *et al*. 2009; Prorocic *et al*. 2011; Gersbach *et al*. 2011; Mercer

---

ICAR- Indian Veterinary Research Institute, Animal Genetics, Bareilly-243 122, Uttar Pradesh, India.

Corresponding Author: Shalu Kumari Pathak, ICAR- Indian Veterinary Research Institute, Animal Genetics, Bareilly-243 122, Uttar Pradesh, India. Email: skpathakvet@gmail.com

How to cite this article: Pathak, S.K., Sonwane, A. and Kumar, S. (2020). Exploration of RNA-Guided Recombinase Target Sites for Hyperactivated Recombinase Beta in Bovine Genome. Indian Journal of Animal Research. 54(9): 1086-1097.

Submitted: 11-07-2019 **Accepted:** 14-08-2019 **Published:** 09-11-2019

et al. 2012). More recently, CRISPR-Cas9 technology has also been modified to base it on recombinase action instead of nuclease action. In this, a hyperactivated Gin (β) recombinase has been fused to the nuclease-null Cas9 or dead Cas9 (dCas9) (Chaikind *et al*. 2016). recCas9 enzyme operates on a minimal pseudo-core recombinase site, a 5- to 6-base pair spacing between the Cas9 and core sites flanked by two guide RNA-specified DNA sequences or Cas9 binding site followed by protospacer adjacent motif (PAM). (Chaikind *et al*. 2016).

Currently, though only one RGR platform based on hyperactivated recombinase Gin (β) is available, more hyperactivated recombinases have been evaluated as part of ZFR platforms. (Gaj *et al*. 2013; Sirk *et al*. 2014). The structure of RGR platform based on hyperactivated recombinase Gin with the degenerate sequence is: CCN (30-31)-AAASSWWSSTTT-N (30-31)-GG (Chaikind *et al*. 2016). RGR target sites for the RGR platforms based on the hyperactivated recombinases other than Gin (β) can be easily discerned by only replacing the central 20 bp pseudo-recognition sequence in the target sequence as described by Chiakind *et al*. (2016). Sirk *et al*. (2014) has reported the core sequence for hyperactivated recombinaseBeta as:NNNV NNR GT WW AC YNN BNNN where N=A, T, C, or G; V=A, C, or G; B=T, C, or G; R=A or G; Y=T or C; W=A or T.

Therefore, in present study a detailed map of target sites for RNA-guided recombinase (RGR) platforms based on hyperactivated recombinase Betawith spacer length five was prepared throughout the bovine genome and genes in context to RGR target sites were identified in entire bovine genome.

MATERIALS AND METHODS

RNA-guided recombinase (RGR) platform operates on a typical recognition or target site comprised of the minimal pseudo-core recombinase site, a 5 to 6-base pair spacer flanking it and whole central region flanked by two guide RNA-specified DNA sequences or Cas9 binding sites followed by protospacer adjacent motifs (PAMs). Therefore, structure of RGR platform based on hyperactivated recombinase Beta with the degenerate sequence will be: CCN(26-27)- NNNV NNR GT WW AC YNN BNNN -N(26-27)-GG.

For this chromosome wise whole genomic sequence data was retrieved from Ensembl (Bos_taurus.UMD3.1.dna. chro-mosome). Search pattern was designed depending on structure of RGR target site which includes degenerate sequence of 20 bp core recombinase recognition sites (NNNVNNRGTWWACYNNBNNN) flanked by spacer (5 bases), guide RNA binding site (25 bases) and PAM sequences at both the ends. Designesd search pattern was as follows:

CC[ATCG][ACG][ATCG][ATCG][AG]GT[AT][AT]AC[TC][ATCG][ATCG][TCG][ATCG]GG.

Genomic sequence data and search pattern was loaded to dreg program of Emboss package individually for each chromosome and target Sites for Rna-guided Recombinase (RGR) were identified throughout the bovine genome. All identified RNA guided recombinase target sites were converted into Fasta format. Fasta sequences were mapped against RefSeq Genome Database (refseq genomes) of *Bos Taurus* (taxid: 9913) for highly similar sequences by using Blast of NCBI. Sequences (Subject sequence) having 100% alignment along with no gap with RGR target site sequences (Query sequence) were selected and features of these sites were identified.

Table1: Chromosome wise number of RGR target sites in bovine genome.

Chromosome no.	Chromosome Size (MB)	RGR target site no.	Chromosome no.	Chromosome Size (MB)	RGR target site no.
1	45.6	22	16	23.6	12
2	39.6	18	17	21.7	12
3	35	19	18	19.1	18
4	34.8	14	19	18.5	16
5	34.9	23	20	20.8	8
6	34.2	18	21	20.6	19
7	32.2	16	22	17.8	16
8	32.6	22	23	15.2	15
9	30.5	9	24	18.2	12
10	30.1	20	25	12.4	7
11	31	19	26	14.9	5
12	26.2	13	27	13	4
13	24.4	14	28	13.4	6
14	24.3	16	29	14.8	10
15	24.5	16	X	42.1	17

RESULTS AND DISCUSSION

In present investigation, 436 RNA-guided recombinase target sites were identified in bovine genomeby using dreg program of Emboss package. RGR target sites were found to be located on all the chromosomes. Chaikind *et al*. (2016) have reported that they searched for appropriate target sites by using Bioconductor, an open-source bioinformatics package using the R statistical programming (Gentleman *et al*. 2004) and identified approximately 450 such loci in the human genome such that location of RGR target sites lies on all the chromosomes of human genome. (Table 1). Further, these sequences were mapped against RefSeq Genome Databaseof *Bos Taurus* (taxid: 9913) for highly similar sequences through Blast to investigate that whether these RGR target sites are present with in any gene or not and it was found that 173 RGR target sites lies in genic region and 263 in intergenic region. RGR target sites lying within genic region has been shown in Table 2.Chaikind *et al*. (2016) also searched ENSEMBL (release 81) to identify which predicted recCas9 target sites fall within annotated genes (Cunningham *et al*. 2015).

In present scenario, there is requirement of improved technology which is capable of introducing targeted modifications in both safe and efficient manner so that genome engineering may reach their full potential in clinical and industrial applications. Most contemporary genome engineering processes are based on the use of targeted nucleases, such as ZFNs, TALENs and CRISPR/Cas9; however, these tools have the potential to introduce potentially toxic off-target double stranded breaks and relies on the host cell machinery to facilitate targeted DNA integration which prevent their use in post-mitotic cells. SSRs, however, offer a potential solution to these problems, particularly for applications of therapeutic gene integration (Crooks *et al*. 2004).

Despite their potential, new approaches for reconfiguring their specificity are needed. Programmable tools for genome engineering based on recombinase action have been devised by fusing these ‘hyperactivated’ recombinases with Zinc-finger proteins and TALEs (Mercer *et al*. 2012; Gaj *et al*. 2013; Gaj *et al*. 2011; Gersbach *et al*. 2011; Sirk *et al*. 2014). Gaj *et al*. (2014) have developed a diverse collection of re-engineered Gin recombinase catalytic domains suitable for the design of ZFRs with custom specificity. They have shown that ZFRs can be assembled to recombine user-defined DNA targets, and that designed ZFRs integrate DNA into endogenous genomic loci. Recently,hyperactivated Gin (β) recombinase has been fused to the nuclease-null Cas9 or dead Cas9 (dCas9). This recCas9 or RNA-guided recombinase (RGR) has been demonstrated to do GE efficiently wherein it precisely deleted a portion from human genome (Chiakind *et al*. 2016). Till now, hyperactivated Beta recombinase has been evaluated as part of ZFR system. Similar to Gin recombinase, hyperactivated Beta recombinase may be fused to the nuclease-null Cas9 or dead Cas9 (dCas9) to develop RNA-

Table 2: Sequences of RNA guided recombinase target sites identified in genic region of bovine genome.

Chr. No.	Start position	End position	RGR Target site sequence	Gene description
1	93795771	93795846	CCTCCATGCTGGTTTCTGTCCTGGCTGTGGGGGTGGTTTACTGCTTGAAGGGGGAGTCTTGGTAGCCTGGCATTGG	multidrug resistance-associated protein 4
1	141817946	141818021	CCTTGTCAACTATGGGAATTTCCGAAACCTTGTTAGTTAACCGATTGACTAACCAATTCTTGAAGTCACCTCTTGG	Down syndrome cell adhesion molecule
1	145858575	145858650	CCCGACCCAGAAATCAGTTGACTGTTAACAAAAGGGTTTACCTGTTCTATTCCTATTAACCATGTGTCTATCTGGG	transient receptor potential cation channel subfamily M m...
1	147676551	147676626	CCCCATGCATGGCGCCAGCACCATCTCAAGCAAGAGTATACCAACCCACGGTCAGTACCTGGGCCGAGGACCAAGG	germinal-center associated nuclear protein
2	14456327	14456402	CCAGAATGGGAGGAGAAAATATATATTCAAAAGGAGTATACCAGTATCAGTGCTGGAATAGGAGTGGGAGTGGAGG	calcium/calmodulin-dependent 3',5'-cyclic nucleotide phos...
2	15723312	15723387	CCATCTGATAAATCTTCCTGAAATGGATGCAACTAGTAAACTTAGTCTTATTTTAAAGACAATCTTAATGAAAAGG	ubiquitin-conjugating enzyme E2 E3
2	73505057	73505132	CCTTTCCAAGCTTCACTCTCATCCAGGTTACAGGAGTTAACCTGCAGGGAGCTGCTGGTAAACAGACCAGCTCGGG	CLIP-associating protein 1 isoform X7
2	103319972	103320047	CCAAAATTAATCTGAAAACTGTTACTGAATTAAGAGTATACCTCCCTTCAGCACGTAAGAATTATAAAGGAAAAGG	BRCA1-associated RING domain protein 1

Table 2: Continue...

Table 2: Continue...

2	106746267	106746342	CCCAACCCGCAGAGGAGAAACTGAGGTCCAGAGAGGTTAACCAACGTGTCCATGGTCACAGACCCAGGCCTCGGGG	tensin-1 isoform X4
2	116910569	116910644	CCAATATAGCAACATTATTTTTAATTGTTTTAAAAGTATACTTTTAAATTATGGAAATTGAGCTTGAAAACATTGG	A-kinase anchor protein SPHKAP isoform X2
2	127702051	127702126	CCTTTTGTTTTCGAGATGCTTTCATATGCCACACAGTATACCCATATTATACCTCTGTTTGCCATGGGTAGGTGGG	platelet-activating factor acetylhydrolase 2, cytoplasmic
2	128151461	128151536	CCATCCCACTGATACTTCTAAAAGATGAAATGATGGTTTACCAAGGATTCTACAAATGCCAGAACTGAAAAGAAGG	macoilin isoform X1
2	133661692	133661767	CCCGCTGCGGCCCAGACCAGTGATTCTAGCTCGAGGTTAACCGAGAAGGCAGACGACTCTGCACATCCCGAGGAGG	MICOS complex subunit MIC10
2	134047454	134047529	CCTTCTTTGGCTCAGCAGTGATTCTCTTGGTCGTGGTAAACCTGTCCAGTTGAAAAACATGGGGGGCGGGGGGTGG	E3 ubiquitin-protein ligase UBR4 isoform X2
3	3098087	3098162	CCACTAGCGCCACCTGCGAAGCCCCTTCCAAGTAGGTATACCCTCAGAATTTAGCACCATTAATTTTCGTTACAGG	uridine-cytidine kinase 2
3	16199645	16199720	CCATTTTACAGATGAAAAGACTGAGGCTCTGAGAGGTAAACTAACGTTACCAAGGTCATACAACCTGATTTCCAGG	interleukin-6 receptor subunit alpha precursor
3	16444069	16444144	CCCTATCCACACAGACTGTACTCTTTTTCCTCCTAGTAAACTCTTGTCTTATTTCACTACTTTCCATCTTTGCAGG	nuclear pore membrane glycoprotein 210-like
3	35826154	35826229	CCCATATATTCGTTTAAGATCTGAGGAATTAGCTAGTAAACCAGGAAAGAGGGAGATACAAAGTCTAGGAAATGGG	guanine nucleotide exchange factor VAV3 isoform X1
3	77401882	77401957	CCACATTGATGCTGTTGTACTCATTTTACACATAAGTAAACTGAGATACAGAAAGCTCTAGTAAAATTGGTCAAGG	protein wntless homolog isoform X1
3	91759276	91759351	CCATTTTGTAGATGAGAAATTTGAGGCACACAGAGGTTAACTACTTGTCTAAATTCACAAGCCTTGTTACAGCAGG	ubiquitin carboxyl-terminal hydrolase 24 isoform X1
3	92445692	92445767	CCCTGCTGGGAGGAGGTGGTTGGCTCCCTCAGTTGGTAAACTGGGCCCTGTGGCCTTGAGTGTTTCGGTCTCTGGG	single-stranded DNA-binding protein 3
3	94647760	94647835	CCCGCTGACCTGAAGGAAAGGAAAACACCCACGTGGTAAACTATCTATGAGGCAAGAACTGCAGGCAGCCTCTAGG	zinc finger FYVE domain-containing protein 9 isoform X2
3	95208971	95209046	CCTCCAATTCTAGTCTAACACCACATGGCTTACAGGTTTACTGGCACAAAGACAGACCAATGAAACAGAGGAGAGG	oxysterol-binding protein-related protein 9 isoform X4
3	97614932	97615007	CCTCTTTCTACCTTGTTCTGGTTTGGAGGAAGCAAGTTAACTTTTGTGAGCTATCTTCTGACCAGTAAAATAAAGG	oxysterol-binding protein-related protein 9 isoform X4
3	113225768	113225843	CCCACCTTTATAAAAAATGATAATAAGAAGCAAGAGTATACCTTTCATTACAAGGAAATAAAAAGGGAGAGGAAGG	ephexin-1 isoform X1
3	114192170	114192245	CCGGAGCACCGACGTGTCCTACGCCGAGAGCGTGAGTTAACCGCGTCCCCTGCAACACGCAGTTAAGGTTCTTTGG	transient receptor potential cation channel subfamily M m...
5	29138839	29138914	CCACTAAATGAAGCTGAAATACCAGAACTTCCTTAGTATACTTCCATGGTAGAGCAGTGGTTCTTCTGACCTGTGG	HIG1 domain family member 1C

Table 2: Continue...

Table 2: Continue...

5	35148869	35148944	CCAAACAGGGATGCCATGACTGACTTGAGCCATAAGTAAACTTTTATTTTATTAAGCCACTGTAATTCTAGGCTGG	anoctamin-6 isoform X2
5	36360030	36360105	CCATCTGGGAAACCTTTTCTTTTATGTGAGGAAGGGTTAACTAGGATGGATAGGGGCCATTTGTCAGGGTAACAGG	transmembrane protein 117 isoform X1
5	45090098	45090173	CCTTCATGGTGGGGTATAAAGGCATTAACAACAGGGTAAACCCTTCAAATATGCAGGCACTGTCACAAAGTCTGGG	carboxypeptidase M isoform X2
5	47650146	47650221	CCAAGCCTACTGACCACAAGTAAAATAGAACCCTGGTAAACCACCAAATAGAGCTGTTGAAGGCTTGATGTGGAGG	glutamate receptor-interacting protein 1 isoform X2
5	48151854	48151929	CCATTTTATAGGGAGGAGTGGTACAATGGAGGGTAGTTTACTATGTTTGCTGAGGAGTTACCCCATTGACAGAAGG	high mobility group protein HMGI-C
5	49989235	49989310	CCAGGTAAGTCCTATTACCTCAGTTTTACAGAGGAGTAAACCGAGGCAACTTGCCCATGGTCACTTGGCCAGCAGG	SLIT-ROBO Rho GTPase-activating protein 1 isoform X1
5	60953594	60953669	CCTTATTATATTTTAAAAGACTGAAATCATACAAAGTATACTTTCTGATCACAATAGGAAAAGGAGTACGTCAAGG	cyclin-dependent kinase 17 isoform X1
5	76273767	76273842	CCCAGAAGTGTCTTCCTCCTCCGAGATCTCTACGAGTTTACTCCTAGAGACTCTTCCCTCTTAGAGAGAAGTCAGG	beta-1,3-N-acetylglucosaminyltransferase manic fringe iso...
5	89694860	89694935	CCTCACATGTGTGACAGCCCACTGCTATCTACCAGGTATACTGATGCTCACACACCAAGTCAGTCCCCATTTTTGG	LOW QUALITY PROTEIN: cGMP-inhibited 3',5'-cyclic phosphod...
5	106726410	106726485	CCTGGGGAAGTGATGCCAAGAGGCAGGAATTGAGGGTTAACTGGGAGGCGGTGGGGGGAGGCCAGGGTTTGGGGGG	EF-hand calcium-binding domain-containing protein 4B
5	110243521	110243596	CCTCCAAACATTCATCAGAGGTGGGGGGGAAACCAGTTAACTGTTAAAGAGAATTCTGGGAAAATGTATAAAAAGG	UPF0193 protein EVG1 isoform 1
5	117785827	117785902	CCCGGGGTGGGATCTTATCAAAAGAGAATCGGCGGGTTTACCACGGGGTGACTGTGCTGGTCAACTGACACTTTGG	cadherin EGF LAG seven-pass G-type receptor 1
5	119876232	119876307	CCGAGGACGGCGCTTACGGGGCCTGCGTGCGCGTGGTAAACTCCTTGAACTCGCTGTCATGGATGGTGAAGTAGGG	protein DENND6B isoform X2
6	13393667	13393742	CCCTGGGGTTGTCACGCTGAAACGGAAAAGCCCCAGTATACCATGAGAACACAGGCATTTGTGACCCCGTTGAAGG	ankyrin-2 isoform X1
6	27565159	27565234	CCTCCCCTGCACCTATAACCTTCTTTCTGGACTCAGTTTACTTTTTTGGGGGGAATCATTTTCCTTTTGGAGAAGG	rap1 GTPase-GDP dissociation stimulator 1 isoform X5
6	30519672	30519747	CCCTTTCAAAAACTAAACAAAGAGCAAAAAAGCCAGTTAACTTTTTTGAGTCTGTTTCCTCATTATTGGAATGAGG	netrin receptor UNC5C isoform X2
6	45259904	45259979	CCCAAGTTTGCTACAAGCTTCAAGAAAAATAGCAGGTAAACTGATTTTCTTTAAGCCCGTTCTTGACTGAACAAGG	peroxisome proliferator-activated receptor gamma coactiva...
6	69384874	69384949	CCTAAAATATTTAAGATCTGGCACTCGTGCAGAGAGTTTACCAACCTCTGCTTTATATTATGTATGGTCGGATAGG	DCN1-like protein 4 isoform X1
6	89371107	89371182	CCAAGTCTGAGCACTGAGTCCACCTCTAAGAAACAGTTAACTTCTCAAAGTCTCAGTTTCATCATCTGTGAAAAGG	A disintegrin and metalloproteinase with thrombospondin m...

Table 2: Continue...

Table 2: Continue...

6	118558216	118558291	CCTGGGGGCTCCCTCCTGCTTCAGCCTTACCAGAAGTTAACCCTTTCATCTCCAAGTCTTCGATACAGGGCCTTGG	VPS10 domain-containing receptor SorCS2 isoform X1
7	19829959	19830034	CCCAGTGGACAGGGGCTCCTCAGGGCAGGCCCAGGGTATACCTCGGTGACCGCATCTACACAGCCAGGGGTCTTGG	lon protease homolog, mitochondrial precursor
7	21368986	21369061	CCTCCCAAAAGGTAGGCGTTCCATTCTACAGGTGAGTAAACTGATCGCTCAAGGTCCCACGGTGAGCAGAGACAGG	zinc finger RNA-binding protein 2 isoform X1
7	45186584	45186659	CCGTCGTGCGCATTCACGATCCCTGCGCTCTGATGGTATACCATGTGTATCTGCGGGACGGTGCACCCCGCGCTGG	rho GTPase-activating protein 45 isoform X2
7	74440057	74440132	CCTAGTTCCTTCTGGGGGATGGGGCATAAAGGTAGGTTTACTGGTACCAGTTTTTCTTCCACATGAAGGAAACTGG	probable phospholipid-transporting ATPase VB isoform X1
9	43990626	43990701	CCACTGTGAAATTACTCTTTCATCCCCTTCCCATAGTAAACTCTTTGTAAGGAAGTCACTGTGTGGAGTGGGGTGG	beta/gamma crystallin domain-containing protein 1 isoform X1
9	69194097	69194172	CCTTTCTTCTACACTCATTCCTGGGTATACACATAGTTTACTTTTTAGCTCTCATTTCTTCAGAAGAGGAGTCAGG	lethal(3)malignant brain tumor-like protein 3 isoform X1
10	7455279	7455354	CCTCAGTCCCCAGGGTTAGAATTTCCTCTGCAGGAGTTAACTGTGTCCAGGAGAGCTTGACCAAAAGCATGGAAGG	synaptic vesicle glycoprotein 2C
10	13340889	13340964	CCAGCCCTGCTTTCTGTTGAGTTTACTCACTGTGGGTTTACTTTGAACTTGCTAGTTCCTGGACAGCTCTGAAAGG	protein zwilch homolog isoform X1
10	35580406	35580481	CCCAGTTCTACCATTTACTATGTGATCACGGGTAAGTTAACTTCCCTGTGCCTCAATTTCACTGTCTGTAAAGTGG	probable G-protein coupled receptor 176
10	35667348	35667423	CCTCACTGCTTGGCTCCTGGGGCTATAACTTGGGAGTTAACCTGCATGATGCCCACCTAAGAAGACACTAAGAAGG	probable G-protein coupled receptor 176
10	52715367	52715442	CCAGAGATCAATTAATTAATCTAAAAAATCAAACAGTAAACCCTCTTATAGAACAAAGGGAAGAAACCCATCTGGG	myocardial zonula adherens protein isoform X3
10	72349477	72349552	CCCAGGTATGGTCAACTCTCAGCTCCTGGTTCTGGGTTAACCTTGATTTAACAAGATTAACAGTTCTTCACCTAGG	reticulon-1 isoform X1
10	79480436	79480511	CCTTGGCTGGGGCTTCTTCCCACATGGCACTAGTGGTAAACCCGCCTGCCATTGCAGGAGATATAAGAATCATGGG	protein FAM71D
10	84435133	84435208	CCTCGTTTATGGACTTGAGCATCAACTACTCATAGGTAAACTTTCAGTCTCCTTGAAAGGACAGCTTATCTCTGGG	regulator of G-protein signaling 6 isoform X8
10	88137324	88137399	CCTTTCATGAGAAGTGTGGAAATAGTCTCTCAAAAGTTTACCAGTGCCACCATGGCTTGAATGGTTGGTTCGTGGG	LOW QUALITY PROTEIN: tubulin polyglutamylase TTLL5
11	207246	207321	CCAACGTTGTCCTGGGCCCACACTGCCTTCAAGTAGTTTACTGCCTCAACTCCGCCCCTATCAGCCAGTGCCCAGG	fibulin-7
11	1430190	1430265	CCCTTGTAGAATTTACTGTCAACTTTGGTTTGGTAGTATACTCATATCCAGCATCACTTCCACCGGGAGAATGCGG	acyl-coenzyme A oxidase-like protein isoform X1
11	3252304	3252379	CCAATCCTAAAGATAGTGAACCTAGACAGGTGGAGGTTAACTCTGCTCACCTGCCCTCACACAGCCAGAAAGCAGG	transmembrane protein 131 isoform X2

Table 2: Continue...

Table 2: Continue...

11	7031219	7031294	CCAAAAAACAATTCAATTGAAGTAAAACTTGGTGAGTAAACTTATCGGAGCCTCTGAAGTTGCTAAGGGCCGAAGG	interleukin-1 receptor-like 2 isoform X1
11	10624491	10624566	CCAGCCTAAACTTGTTCAAGCCAACAGAGAAACGAGTTAACCCAGGGGGAAGGTGCAGCAAAGGGAGTCTGTCTGG	methylcytosine dioxygenase TET3 isoform X1
11	70772164	70772239	CCTGTGCTGGAAACAGGCTGAGAAGGAAGGGCACAGTTAACCATGAAGCCACCTGGTCAGGCCCAAGGGTGGTGGG	LOW QUALITY PROTEIN: TOG array regulator of axonemal micr...
11	104173969	104174044	CCCCACCTTAATGAGAGGAGACACCCTGAGCCTCAGTTTACCACGTGTGAAACACACTCCTGCGTTGGAGGTGAGG	protein FAM69B isoform X3
12	13464313	13464388	CCTTAAGACCTGAGGGTGATGAAGACTATGGCTGGGTATACCCACCCAGGCACCTCTCTCTCCACCTCTGGCTCGG	ecto-NOX disulfide-thiol exchanger 1 isoform X2
12	24165065	24165140	CCTGTACCACTTCCTACCAGTATGATTTGGGACAAGTTAACCTATGCCTCAACATTCTTCTCTGTAAGATGATAGG	short transient receptor potential channel 4 isoform X1
12	79734136	79734211	CCACGTTCCTGGTAAATGCTTTCATCTGGCCAAGGGTTAACCTCGCCAAAGCCTGTGTCCACCAGTTGCAAGCAGG	dedicator of cytokinesis protein 9 isoform X7
13	37999535	37999610	CCGTACCCTAGATACACAGACGACTGGTTTAACAGGTAAACTGGGCCTCTGGGGATGTCTGGGACATCAGGAGTGG	neuroendocrine convertase 2 isoform X1
13	39426183	39426258	CCCGTGGGTGTTGTATGTTTGCATCCAAGAGCAAAGTATACTCCTTGCTCCAGAGGGAAGCCTCTTGTGCTTCTGG	sodium/potassium/calcium exchanger 3
13	53868309	53868384	CCATCGAGTCGCAAGAGTTGGACACGACTTAGTAAGTAAACCACCACCACCCAAAATCTCCTTGAATTAACCATGG	tyrosine-protein phosphatase non-receptor type substrate ...
13	61266273	61266348	CCAGTTTACAGATGAGAAAATTGAGGCTCAGAGAGGTAAACCCACTTGGTCAAGGTTACAGATATCAGTAAGGGGG	uncharacterized protein C20orf96 homolog isoform X3
13	76339605	76339680	CCCTGAGTGGGGACCACGCCATCATTTACAGATGGGTAAACTGAGGCACAGCAGCAGCAGCGTGACTTGTGTAGGG	eyes absent homolog 2 isoform X1
14	26815359	26815434	CCCCATTGTCAACTGTACTGCTCTCGAGAACCCTGGTTTACCCTGGTTGAGTTTGAAGCTTTGGTTGTTTTCATGG	thymocyte selection-associated high mobility group box pr...
14	33380982	33381057	CCTAACTTCTAGCTTCTGTTATCGTGTAATGATAGGTAAACCAATGGAAATAAGTTGCTGGTAGACAACAAGGAGG	brefeldin A-inhibited guanine nucleotide-exchange protein 1
14	59359973	59360048	CCAGGTACTATTCAGACCACTAGGGAGAGAGAAAAGTATACTGTCTGTGGCTTCAAGAAAACTAGAATTTGGTAGG	angiopoietin-1 precursor
14	67200088	67200163	CCAGTATTTTAGAAAGGGATCTGAGACTATAAGAGGTTAACTAACTGGCCCAAAAAACAAAGTCCAAAGAAGAAGG	vacuolar protein sorting-associated protein 13B isoform X5
14	76924502	76924577	CCTTGAGCCATGGAGGAACTGAGATCAGAGCAGCAGTTTACCAAGTTGGAGAAGCACCCACAGCCAGAGATTCTGG	matrix metalloproteinase-16
14	80135070	80135145	CCAGTCAAAGCAGAGGAAAAGGAAAAGACAAAGCAGTATACTATTTGAAGAGTTTTGTTTGGTTTATGAAGAATGG	RNA-binding Raly-like protein isoform X5
15	6882511	6882586	CCCATATGGCATAATTTTCAGATCATTTTGTCCTGGTTTACCTTCGCCAGAGGCTCACTTGAGTGGGGTGTTTGGG	transcriptional coactivator YAP1 isoform X8

Table 2: Continue...

Table 2: Continue...

15	8408474	8408549	CCTCATGTCTCTTTTATCTAAAAAAACCAGCCCAAGTTAACTGCCTACTGTATGCTGGGGAATTTTTAAATCAAGG	rho GTPase-activating protein 42 isoform X1
15	27970862	27970937	CCAAGCAAGTCTGATCTACCAGGGAGCCTGAAGAGGTTAACCACGTGTCTTGGCCACTGTATCAGATCTCATCTGG	serine/threonine-protein kinase SIK3 isoform X4
15	28737997	28738072	CCCCTTTTATAGATGAATAATTGAGGCTCAGAAGGGTTAACCCAGGACACAGAACCCAGTCAGTGGGGGTGCTGGG	Down syndrome cell adhesion molecule-like protein 1
15	35060215	35060290	CCAGGGACCCGGAGGTGGGAAGGGAAGGGGCAGGAGTTAACTGTCCGCCTTGGTCTTGATCTCTTCGAGCGTCTGG	secretion-regulating guanine nucleotide exchange factor
15	51814932	51815007	CCAGGGCCCACCTGGTCTGCTCCTGCACAGAAAGGGTTAACTTTCAACATTCGATTATCAGGGTACTGGGATGCGG	ribonucleoside-diphosphate reductase large subunit
16	29654138	29654213	CCCTTCCTCTCTTCCTATGAACATTTTTCCTGCAAGTAAACTTGGGTTGGATAGCCGCCTCCTTTAGGGCATGGGG	CSC1-like protein 1 isoform X1
16	30990105	30990180	CCATTGCCTTCTCCAAAGAAATTCCCTGGTTCCAGGTAAACCTGGATGGTGGGTTACCTTACTGGGAGTCCACGGG	serine/threonine-protein kinase MRCK alpha isoform X4
16	53816375	53816450	CCTCCAGCCATGTCTTCAAGCTCACTTTTTTGAAGGTTTACTTTGTATGAAGCACCGTTCAGGGCTTCCCTGGTGG	kazrin isoform X1
16	71139373	71139448	CCAGGGCAGGGATGGGCCTTTCTAGATACAAAGTGGTAAACCCAGAGAAGAGAGCCAGGTGCCTGACATGGCCCGG	protein phosphatase 1 regulatory subunit 12B isoform X10
16	77735516	77735591	CCAGCAAACCCTTGGGGTGCCTACTTCGTGTAAGAGTTAACCATGTTAATGCTGGAGACTCTAATGTTTCCCCTGG	membrane cofactor protein
17	8890217	8890292	CCAAAGCCAGGATACCAAGACTCTCCCTTCAACCAGTAAACCACCCCCACAGCCAGGACAGAAACCAGCATGGAGG	LOW QUALITY PROTEIN: multidrug resistance-associated prot...
17	48795294	48795369	CCTTTTCTCAGTAGCTGGGTGACCCTGGAAAGATGGTTTACCCTCCTTGACCCCAAACTCCTCATCTACCAACTGG	CSC1-like protein 1 isoform X1
17	51600154	51600229	CCAGTCTACAGGTAAGTCATGATTTATTCATAAAGGTTAACCTTGGTTAATGTCTTAGAAAATACAGTCTATTTGG	serine/threonine-protein kinase MRCK alpha isoform X4
17	62028677	62028752	CCACTGCTGTCGTGGCCATGGAAGCATGTGTCAAGGTAAACTCTCTATTAGTCTGGATCTCTGAAAGATTTCAAGG	kazrin isoform X1
17	65012948	65013023	CCTGGTAGCCCTCCCTTTTATCGACTCTATCCTCAGTTTACTCAGATGTCAAAAGTAACAGAGCCTACCTTGTAGG	protein phosphatase 1 regulatory subunit 12B isoform X10
17	66189351	66189426	CCTGGCATTGAACACACTTGGAAACTAACTTAGCAGTTTACCTATTTATTTGCACATTTCCAATCTTCCTGCTAGG	membrane cofactor protein
18	8160154	8160229	CCATTTCAACTGCTACTTTCTCTCTGAAGAAGAGGGTTTACTGAGCTCCTTGCTGATTGGGACTGTCTTGCAGGGG	C-Maf-inducing protein isoform X1
18	39365534	39365609	CCCTTTTAAAAGAAGCAGGGGCAGGGAAGAGACTGGTAAACCATCTTAGTGAATTGCCTACTTGTTGTGTTGCAGG	zinc finger protein 821 isoform X1
18	46056063	46056138	CCAGCCCCAAGCCAGGCCTTGGCACCAAGCCACGAGTTAACTGCCCTCTCTGTCTGAGCTTCCAACCAAGTGGGGG	phospholemman isoform X1

Table 2: Continue...

Table 2: Continue...

18	46988385	46988460	CCTTTTTCCCTTCTCAGGAATCAGCCATGTTCCTGGTTAACTCGTTCTTGAAGGGTGGCGGAGGCGGCGGAGGAGG	calpain small subunit 1
18	53655555	53655630	CCCGTTGCCCTCACCCCAGGTTCTGATTCTGCCCAGTTTACCCTCTGGCCAAGCCCTCAAATTCCTTTCACGCCGG	echinoderm microtubule-associated protein-like 2 isoform X2
18	56670186	56670261	CCCGGTCTCAGCTCCTGGGTCCCGGCAGTTGAGGGGTTAACTCCCAGCTGTAGAGAGAGGCGGTAACTGGAGCTGG	bifunctional polynucleotide phosphatase/kinase
18	56960349	56960424	CCGCCGCAGTGCGCAGCTGGACAGCGTCAGGATTGGTTTACCCCCTTGTCCCAAGATGGTTCAAAAGCAGAAGGGG	LOW QUALITY PROTEIN: potassium voltage-gated channel subf...
19	11302922	11302997	CCCTTTAAATAGGTTTATGATTTTAAGACAAGTCAGTTAACTTCTTTGTGTCTTGAATTCCCTACCAGCAAAATGG	mediator of RNA polymerase II transcription subunit 13
19	27933197	27933272	CCCCCTACTCCAGAACAACACAAAAGGTAGTGGGGGTAAACTATGTCTTCCTTGCCTACCTTCTGCCCTGAAAGGG	fragile X mental retardation syndrome-related protein 2 i...
19	31345216	31345291	CCCCTGCTGGGCACTTTACTTTGTGCTTTACATAAGTTAACTCATTTGATCTTCATTAGAGTCCACTGAGGATGGG	dual specificity mitogen-activated protein kinase kinase 4
19	32773833	32773908	CCAAGCAGAGACTGTTTGAATCAGATACCCCCTCAGTAAACCAGCTACTGTGTCCAGGTGTGGGGTAAAGCACGGG	protoheme IX farnesyltransferase, mitochondrial
19	42483661	42483736	CCGGTGAGTTCTTAGACCGATTGGCCTATTAGATAGTTAACTGCCACCACATAGCAGGCTGTCCCTGAGGGACAGG	keratin, type I cytoskeletal 42 isoform X3
19	44702012	44702087	CCAGCGGACCTGGGGGCTCAGACCTCCAGTGGTCAGTTTACCTCTGTCCAGCGAGCTCCCCATGAGTGCAAGGAGG	band 3 anion transport protein
19	46765808	46765883	CCTAGATTAATACAGTCGGCCCTCTTTAACTGTGGGTTTACCACCTAAGGTTGGTTGAATCTGCAGGTACGAAGGG	KAT8 regulatory NSL complex subunit 1 isoform X2
19	50541960	50542035	CCTCAGCAGAGATTCATAGGGTCTTAGTAATCCTGGTAAACCATGGGGTACAGTTCTAGGTACCCCAGATCTGAGG	fructosamine-3-kinase
19	54724936	54725011	CCATCAAACTGTACACACAGAGGGTGAACTCACAGGTAAACTGTGGACTTTAGTTAATAATAACTATCAATACTGG	trinucleotide repeat-containing gene 6C protein isoform X8
20	23696620	23696695	CCATGTTGACATAAGTGCTGTGCTGCCTTTTCCCAGTTTACTCCCAGTCAGCACAAGGCAGTAAGCTTTGCTAGGG	phospholipid phosphatase 1 isoform X1
21	7676809	7676884	CCCCATTCTCCCTGACCCCACCCTACACTCACCAAGTAAACCCCGGTGGAAGAGAGAAAATGCTTGGAGCCTGGGG	tetratricopeptide repeat protein 23
21	12949303	12949378	CCAACCAAGAGACTCTTGCAAACCTGGTTAGCACAGTTAACCTCTTCTCGTAGGTCCCCATAAACGTTCTGCTTGG	multiple C2 and transmembrane domain-containing protein 2...
21	23337185	23337260	CCCTACCCCTTCTGTATTGACAGAAAGTCTGAGAAGTAAACTGTCAACATGGGCAAATGCTCAGTAAGCTGAAAGG	cytoplasmic polyadenylation element-binding protein 1 iso...
21	27930810	27930885	CCATGTTCCAAGGTGGGGCTCTGAGCTCCAGAGAGGTTAACTAATGCTTTAACATCACACAGCCGGAGGCCAGAGG	fanconi-associated nuclease 1
21	29356531	29356606	CCTCTCTTGCTCCCTCAGGCCCTGCCATGGGCTGGGTAAACCACGGGGACTGTGCTGCTGGGGGCTGTTGAGGTGG	TM2 domain-containing protein 3 precursor

Table 2: Continue...

Table 2: Continue...

21	43247730	43247805	CCATGAGGAAATGCAGGATTGAGTTTGCAAGGAAAGTTAACTGTTGCTCTCTAAAATGAGATCTTCCCAACCCAGG	A-kinase anchor protein 6 isoform X1
21	45781237	45781312	CCTAAAATAGTTACTCTCTCTCCTTTGTGGGAAAAGTTTACTCATTCCTGTTTAAGCTGCTATACCTAAAGCTTGG	protein FAM177A1 isoform X1
21	66221798	66221873	CCCCGCTTGACCCATGAGGAGAGAGGCACAGAGGGGTTAACCGGCCAGCCCAAGGTAACACAGCAGCCAAACTGGG	coiled-coil domain-containing protein 85C isoform X3
22	4672266	4672341	CCCCCAGCCACCATTTTCCCCTCTGGTAACCATGAGTTTACTAAGTCTGTGAGTCTGTTTCTGCTGTATAAATAGG	RNA-binding motif, single-stranded-interacting protein 3 ...
22	7314886	7314961	CCCTCTGCATTGGGAGCACATAGTCTTAGCCGCTGGTTTACCAGGGAAGTCCCAGATTTTATATTTTGATGAGAGG	E3 ubiquitin-protein ligase TRIM71
22	14152948	14153023	CCACGAAATTAGATTTAAACTTGAATGATGCATGGGTTTACCAGCCATATCAAGGTGAGGGGAAAGCATTTCAGGG	serine/threonine-protein kinase ULK4 isoform X4
22	48669078	48669153	CCTTCCAGGCTCAGCTGCCTGGTTCTGGCCCAGGGGTATACCCACATACCTTGCCACCTTTGGGATTGCGTGTGGG	inter-alpha-trypsin inhibitor heavy chain H1 isoform X1
22	54225299	54225374	CCTTGAAGCGCTACCTACACAGAACATTGCTATCAGTTTACCCTCAGGAGAGCTTGAGAGAGCTTCAGGGAAAAGG	phosphatidylinositide phosphatase SAC1 isoform 1x
22	61243856	61243931	CCCCGGTGTCCAGACCGTCGTCCGCCCAGAAGATGGTAAACCCCCCTTTCTTGTCCCCGTGGATGAGGGTCCTAGG	cytosolic 10-formyltetrahydrofolate dehydrogenase
23	353887	353962	CCTACTCAAAATCTACAATAAAATTGAAGATGAAGGTAAACTCTCAAACACATCATGATCAAGATTTATTCCAGGG	KH domain-containing, RNA-binding, signal transduction-as...
23	3670209	3670284	CCGTTTCTTTAAAATGAGGATGGTTATACAAGAAAGTTAACTTCCTTTTATTTTGAGTTTATGGTTTGACATTGGG	dystonin isoform X4
23	13788425	13788500	CCTGGGTCTTCTTACAGCCTGACACTCTTGTCTCAGTAAACCATTTGGTGGGGCCTCATCCCTCCTGAGGCGAGGG	disheveled-associated activator of morphogenesis 2 isofor...
23	30969046	30969121	CCATTTGGATTCTTTCTGATTTGGGCATATTATAAGTAAACCTTCTATAAACAATCTTGTACAAGTCTTTTGGTGG	zinc finger protein 184
23	31069021	31069096	CCATTTTACACACGAGGAAACTAAGGTGCAGAAGGGTTTACTCTTGCCCAAGTTCACACAGCGGAGGTGGAGCAGG	thymus-specific serine protease
23	43370255	43370330	CCTGCCCACCTCTGCTGGCCTGCGCTTGGCAGGAAGTATACCAGGGCAAGCCAGGATGCACACTCCTACGTCTTGG	phosphatase and actin regulator 1 isoform X4
24	50013198	50013273	CCCAATCCCAGAAGTTACATTTTATGGAATGGAGAGTAAACCTACAGAGCACGCCAGCCAGAGTAGATGAGCCAGG	unconventional myosin-Vb isoform X1
25	277398	277473	CCTCTCACTTTTCCCTTCATTTCTTTTTTCTGTTGGTTTACTTCTGTAGTTATTTGGGAAGCTCAGAAGCATCAGG	protein FAM234A isoform X2
25	25584730	25584805	CCCAGTCTCAGCGACACATCTTCCTTCCAGGCCAAGTTAACCTCGAAACAAGTCAGAATGCCCACGGGCTGCCAGG	protein KIAA0556 homolog isoform X2
25	36807599	36807674	CCTCAAAGAAGGAAGAAAACAATTCCTTCATGAAAGTATACTTTTTAGTTTGAGAGGTTGGATGTGGTAAACCCGG	NXPE family member 3-like isoform X2

Table 2: Continue...

Table 2: Continue...

25	38260645	38260720	CCTGGGAGCAGCTGTGGACTAGACCTCACCCACCAGTTTACTTCCTCTCGTATCTTTACAACACACCACACGGTGG	brain-specific angiogenesis inhibitor 1-associated protei...
26	33099863	33099938	CCCCTAGTTACTGTAGTATGACCTTGGGCAAGTTAGTTAACTCCTCTTTGTCTCCACCTACGAATTTGTTCAATGG	beta-tectorin precursor
26	33568200	33568275	CCTCCCTTGTCAGCCCTGCTCTATTTTGGTAACAGGTAAACCCACATCCTGAATTGGGCTTCCCTGGTGGCTCAGG	vesicle transport through interaction with t-SNAREs homol...
27	42208646	42208721	CCCCGCAACAGCCACGAGGACCAGCACCACCAAAAGTAAACTTATACACACCTGGAAGAACAAAGCGGGCGAGAGG	ubiquitin-conjugating enzyme E2 E2 isoform X1
28	15861545	15861620	CCTTCATTGTATATATGGTAATTCAAATTAGAAAAGTTAACTTTTTTAAAGTTCCTTATCTCTTTGTGTCTTAAGG	ankyrin-3 isoform X4
28	18406934	18407009	CCAAACCCCTAGTCACTGCCATAGTAACTCCCATAGTAAACTACGCTCTGCTTCTAACAGTCTGTCTATTTGAGGG	rhotekin-2 isoform X1
28	27049954	27050029	CCCTTGTCCGCTTGATCATGGTTGGCTACCGACAGGTAAACCTCTTTGTCAGCAGGCAGGATTGGGAGGGAGATGG	A disintegrin and metalloproteinase with thrombospondin m...
28	41019694	41019769	CCAATCTTGCATGATGGAGGAGAAAGGTATTCCAGGTATACCAGGAATACCAAATGATCTCCATAATCATTTGGGG	glutamate receptor ionotropic, delta-1 isoform X2
29	18975811	18975886	CCTACCAAGATTGTTGTGTGAACTTGGAAATGTTAGTTAACCTGTCTGGGCTTTAGTTTGTTTATTTGTAAATTGG	glycerophosphodiester phosphodiesterase domain-containing...
29	34870150	34870225	CCTGCCTCCCTCCATCCCTGCCTTCCTTTGTGCAAGTTTACTCTGCAGGGAGTGAGCTCCTCAGACTTCAGGATGG	opioid-binding protein/cell adhesion molecule isoform X4
29	35821107	35821182	CCTAATTTTTCCCATTTTGCAGGGCAAATACCAAAGTTTACTTGGTCTGTCTACAGCATTGAAAGTGCAACTGGGG	neurotrimin isoform X4
29	48622926	48623001	CCCTGGGTTGCGAAGACCCCTTGGAGGAGGAAATGGTTAACCCACTCCAGTACTCTTGTACTGGAAAATCCCACGG	SH3 and multiple ankyrin repeat domains protein 2 isoform X4
X	5091402	5091477	CCAAGTAAATGGGCAAGTACTTCAAATTAACACAAGTAAACTACTGGCTGAGAGCCATGGACCATGGGCCATCAGG	homeobox protein ESX1 isoform X1
X	7195635	7195710	CCAGCATTTCAGTCTTCATTCTCTCCAGAGCAAAGGTTAACTAGCCCATTCCATAAACATGGAATCACTACCAAGG	glutamate receptor 3 isoform X1
X	32894211	32894286	CCTAGGGTCAACTTGACACCTACGTGAGAGACCCAGTTTACTCTGACTCTGTAGCCACTTCAGGAGGACCCCATGG	LOW QUALITY PROTEIN: protein CXorf40A
X	40414096	40414171	CCACTGTGTTTTGTTACTTTCCCACTGTGGGCTTAGTTTACTTCCCACCATCCTCGGGTGATGGCAGCCCAGTGGG	rab GDP dissociation inhibitor alpha
X	48137994	48138069	CCCAGTAGTAGTATTCATATCTGTGCACGGTGGTGGTTTACTTGCTAAGTTGTGTTCAACTCTTGCGACCCCATGG	protein diaphanous homolog 2
X	57137528	57137603	CCAGAAATCCTTGATGTGGAGAATTGAATGAAGAAGTATACTTGTGTAATGAGAAATATACTTAATGCAAGTGTGG	protein BEX4 isoform X1

guide recombinase platform based on hyperactivated recombinase Beta.

CONCLUSION

Implementation of RGR platform for genome modification requires a detailed map of RGR target sites based on hyperactivated recombinase Betathroughout the bovine genome. RGR target sites are located widely in genome andthese RGR target sites can be utilized for targeted manipulation of DNA *viz*. integration, deletion and inversion. The present map envisages the complete information regarding location of RGR target sites in bovine genome. As site specific recombinase mediated genome editing is more safer alternative of genome modification, this information may be utilized for editing of genome by using RNA-guided recombinase platform.

ACKNOWLEDGEMENT

This work was supported by the ICMR through fellowship (Reference: 3/1/3/JRF-2015/HRD).

REFERENCES

Akopian, A., He, J., Boocock, M.R. and Stark, W.M. (2003). Chimeric recombinases with designed DNA sequence recognition. Proceedings of the National Academy of Sciences of the United States of America. 100: 8688-8691.

Chaikind, B., Bessen L.J., Thompson, D.B., Hu, J.H. and Liu, R.D. (2016). A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. Nucleic Acids Research. 44: 9758-9770.

Chalberg, T.W., Portlock, J.L., Olivares, E.C., Thyagarajan, B., Kirby, P.J., Hillman, R.T., Hoelters, J. and Calos, M.P. (2006). Integration specificity of phage phiC31 integrase in the human genome. Journal of Moecular Biology. 357: 28-48.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Research. 14(6): 1188-90.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., *et al*. (2015). Nucleic Acids Research. 43: D662-D669.

Gaj, T., Mercer, A.C., Gersbach, C.A., Gordley, R.M. and Barbas, C.F. (2011). Structure-guided reprogramming of serine recombinase DNA sequence specificity. Proceedings of the National Academy of Sciences of the United States of America. 108: 498-503.

Gaj, T., Mercer, A.C., Sirk, S.J., Smith, H.L. and Barbas, C.F. 3rd. (2013). A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic Acids Research. 41: 3937-3946.

Gaj T., Sirk, S.J. and Barbas, C.F. III (2014). Expanding the scope of site-specific recombinases for genetic and metabolic engineering. Biotechnology and Bioengineering. 111: 1-15.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al*. (2004). Bioconductor: Open software development for computational biology and bioinformatics. Genome Biology. 5: R80.

Gersbach, C.A., Gaj, T., Gordley, R.M., Mercer, A.C. and Barbas, C.F. (2011). Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucleic Acids Research. 39: 7868-7878.

Gordley, R.M., Gersbach, C.A. and Barbas, C.F. (2009). Synthesis of programmable integrases. Proceedings of the National Academy of Sciences of the United States of America. 106: 5053-5058.

Grainge, I. and Jayaram, M. (1999). The integrase family of recombinase: Organization and function of the active site. Molecular Microbiology. 33(3): 449-456.

Grindley, N.D.F., Whiteson, K.L. and Rice, P.A. (2006). Mechanisms of site-specific recombination. Annual Review of Biochemistry. 75: 567-605.

Mercer, A.C., Gaj, T., Fuller, R.P. and Barbas, C.F. (2012). Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic Acids Research. 40: 11163-11172.

Prorocic, M.M., Wenlong, D., Olorunniji, F.J., Akopian, A., Schloetel, J.G., Hannigan, A., McPherson, A.L. and Stark, W.M. (2011). Zinc-finger recombinase activities *in vitro*. Nucleic Acids Research. 39: 9316-9328.

Sancar, A., Lindsey-Boltz, L.A., Unsal-Kacmaz, K. and Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annual Review of Biochemistry. 73: 39-85.

Sarkar, I., Hauber, I., Hauber, J. and Buchholz, F. (2007). HIV-1 proviral DNA excision using an evolved recombinase. Science. 316: 1912-1915.

Sirk, S.J., Gaj, T., Jonsson, A., Mercer, A.C. and Barbas, C.F. (2014). Expanding the zinc-finger recombinase repertoire: Directed evolution and mutational analysis of serine recombinase specificity determinants. Nucleic Acids Research. 42: 4755-4766.

Smith, M.C. and Thorpe, H.M. (2002). Diversity in the serine recombinases. Mol. Microbiol. 44: 299-307.

Thyagarajan, B., Guimaraes, M.J., Groth, A.C. and Calos, M.P. (2000). Mammalian genomes contain active recombinase recognition sites. Gene. 244: 47-54.