The Pharma Innovation Journal 2023; 12(12): 4092-4102

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(12): 4092-4102

www.thepharmajournal.com
Received: 17-09-2023
Accepted: 20-10-2023

Mitek Tarang
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Subodh Kumar
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Arvind Sonwane
Division of Genetics and Biotechnology, Central Institute of Fisheries Education, Versova, Mumbai, Maharashtra, India

Amrutham Sandeep
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Insha Zafir
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Shahista Sarin Lodhi
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Corresponding Author:
Mitek Tarang
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

RNA-guided hyper activated Beta recombinase target sites identification for genome editing in Buffalo

Mitek Tarang, Subodh Kumar, Arvind Sonwane, Amrutham Sandeep, Insha Zafir and Shahista Sarin Lodhi

Abstract

RNA-guided recombinase (RGR) is a fusion protein composed of nuclease-null or dead Cas9 (dCas9) and ‘hyperactivated’ recombinase. RNA-guided recombinases platforms are potentially valuable tools for basic research and genome editing. The platform recombines a typical target site that comprises of a degenerate recombinase site, flanked by a 5 to 6-base pair spacer sequences 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 (PAMs). Here we have constructed a detailed map of target sites for RNA-guided recombinase platforms based on hyperactivated recombinase Beta throughout the buffalo genome. Firstly, Chromosome wise sequence data of buffalo genome was retrieved from NCBI. Then we designed search pattern for recombinase Beta with spacer length six. RGR target sites were located using this search pattern through Fuzznuc program of Emboss package. The result reveals 426 potential RGR target sites in buffalo genome for recombinase Beta with spacer length six. Out of these 207 RGR target sites lie in genic region and 263 RGR target sites in intergenic region. This work represent a step towards bringing genetic improvement in buffalo and livestock breeding as these RGR target sites provide potential of being utilized for specific genomic integration, deletion or inversion.

Keywords: Buffalo genome, RNA-guided recombinase target sites, genome editing, hyperactivated recombinase, fuzznuc, emboss

Introduction

Efficient, precise and programmable genome editing remains a longstanding goal of researchers across the globe. In the recent years, advancement in genome engineering is reshaping fundamental biological research and biotechnology by making it possible for researchers to introduce custom changes into genomes of any organism. The majority of DNA modification strategies involves the formation of DNA double-stranded breaks (DSBs) or paired single-stranded DNA breaks, which are then processed through endogenous non homologous end joining (NHEJ) or homology directed repair (HDR) (Sander *et al.* 2014, Brookhouser *et al.* 2017) [25, 3]. Site-specific endonucleases based genome editing tools such as Zinc Finger Nucleases (ZFN), Transcription activator-like effector nucleases (TALEN) and Clustered Regularly Interspersed Short Palindromic Repeats -associated protein 9 have emerged as powerful and versatile tools that have drastically improved the ease of targeted DNA modifications. However, it has become apparent that these customizable nucleases are constrained by variety of factors including potentially mutagenic off-target effects (Gabriel *et al.* 2011, Fu *et al.* 2013) [9, 8] and their reliance upon cellular repair machinery for editing after formation of DNA DSBs and it may function erratically at some occasions (Orthwein *et al.* 2015, Ihry *et al.* 2018, Haapaniemi *et al.* 2018) [22, 18, 17].

As a valid alternative site specific recombinases are also powerful tools for genome engineering. They have the enzymatic machinery to perform transient DNA cleavage, strand exchange, and religation without the use of high energy cofactors, or DNA replication and also these processes does not depend upon the cellular DSB repair machinery (Grindley *et al.* 2006) [16]. The super family of recombinase can be subdivided into two basic categories: i) tyrosine (Flp, Cre, and λ integrase) (Grainge and Jayaram 1999) [15] and ii) serine recombinases (Hin invertase, sin resolvase and Gin invertase) (Smith and Thorpe 2002) [27].

Tyrosine and serine recombinases such as Cre, Flp andψC31 integrasehave been widely used to catalyze the recombination of exogenous DNA into model organisms (Turan *et al.* 2013, Gaj *et al.* 2004) [30, 13]. However, the use ofsite specific recombinases has been limited by their intrinsic, non-programmable DNA sequence specificity.

Recombination using these enzymes at endogenous DNA sequences only occurs at ‘pseudo-sites’ that resemble the recombinase’s natural DNA recognition sequence or at genomic sequences for which the recombinase has been experimentally evolved (Karpinski *et al.* 2016, Buchholz *et al.* 2001, Thyagarajan *et al.* 2001) [19, 4, 29].

To circumvent this, recombinases have been retargeted by fusing catalytic-domains to zinc finger or transcription activator–like DNA-binding domains (Akopian *et al.* 2003, Gersbach *et al.* 2011, Mercer *et al.* 2012) [1, 14, 21]. However these techniques require complex addition of heterologous DNA-binding domains to re-target the sequences of interest. More recently, CRISPR-Cas9 technology has also been modified to base it on recombinase action instead of nuclease action. This recombinase–Cas9 (recCas9) was created via fusion of dCas9 with hyperactivated Gin invertase (Chaikind *et al.* 2016) [6]. Another RNA-guided recombinase (RGR) called integrase Cas9 (iCas9) was also generated which was developed through the fusion of dCas9 with the catalytic domain of Tn3 resolvase (Standage-Beier *et al.* 2019) [28].

Currently, these few RGR platform based on hyperactivated recombinase are available, morehyperactivated recombinases have been evaluated as part of Zinc Finger Recombinase platforms (Gaj *et al.* 2013, Sirk *et al.* 2014) [11, 26]. 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) [6]. 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) [6]. Sirk *et al.* (2014) [26] has reported the core sequence for hyperactivated recombinase Beta 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.

Here, in present study we have prepared a detailed map of target sites for RNA-guided recombinase platforms throughout the buffalo genome and we have also reported genes in context to RGR target sites in entire buffalo genome. Our RGR platform was designed to base on hyperactivated recombinase Beta with spacer length six.

Material and Methods

RNA-guided recombinase platform operates on a typical recognition sites that comprised of the minimalpseudo-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. 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. 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.

To accomplish our objective firstly, chromosome wise whole genomic sequence data was retrieved from NCBI (*Bubalus bubalis UOA_WB_1*). 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 (6 bases), guide RNA binding site (25 bases) and PAM sequences at both the ends. Designed search pattern was as follows:

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

Searching for appropriate target sites was done using Fuzznuc program of Emboss package. Genomic sequence data and search pattern was loaded to it individually for each chromosome and target Sites for RNA-guided Recombinase were identified throughout the buffalo genome. All identified RNA guided recombinase target sites were converted into Fasta format. By using Blast of NCBI (National Center for Biotechnology Information),Fasta sequences were mapped against RefSeq Genome Database (Refseq genomes) of Bubalus bubalis (Taxid: 89462) for highly similar sequences. 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.

Results and Discussion

Beta recombinase is a member of resolvase or invertase family under Serine recombinase category and it is obtained from plasmid *Streptococcus pyogenes* “PSM19035” (Rojo and Alonso 1994) [24]. It requires a host-encoded accessory factor such as HU or eukaryotic HMG1 proteins, Hbsu (Alonso *et al.* 1995) [2] and recognize a 90-bp specific sequence, six with only two binding sites (Canosa *et al.* 1996) [5]. Mutant variations of Beta recombinases having the ability to recognize target sites apart from their earlier targets have also been isolated (Sirk *et al.* 2014) [26]. The core target sequence of wild type beta recombinase is given as: CAATAGAGTATACTTATTTC and the core sequence for hyperactivated recombinase Beta is: 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 (Sirk *et al.* 2014) [26]. The structure of RGR platform based on hyperactivated recombinase Gin with the degenerate sequence: CCN (30- 31)-AAASSWWSSTTT-N (30-31)-GG, where W is A or T, S is G or C, and N is any nucleotide was used as search pattern to identify potential target sites in human genome by Chaikind *et al.* 2016 [6]. In the present investigation to identify potential target sites in buffalo genome we used previous finding that characterized hyperactivated recombinase Beta (Sirk *et al.* 2014) [26]. Using this information, we searched the buffalo genome for sites that contained CCN (26-27)- NNNV NNR GT WW AC YNN BNNN -N(26-27)-GG, 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. The N(26-27) includes the N of the NGG protospacer adjacent motif (PAM), the 20-base pair Cas9 binding site, a 6-base pair spacing between the Cas9 and the core target sites.

Scanning search pattern for the entire genome using Fuzznuc program of Emboss package reveals 426 potential RNA-guided recombinase target sites in buffalo genomeand the presence of target sites were found on all the chromosomes. Chromosome wise number of RGR target sites in buffalo genome is listed in Table 1. The highest number of target sites were found to be located on chromosome 3 (48 target sites) and the lowest number of target sites were found to be located on chromosome 16 as well as 24 (7 target sites each). On X-chromosome 21 RGR targets sites were found to be located. There is a report of identification of approximately 450 potential loci in the human genome of RNA-guided recombinase based on Gin recombinase by Chaikind *et al*. 2016, such that location of RGR target sites lies on all the chromosomes of human genome. Chaikind *et al*. (2016) [6] have done the target site searching with the help of Bioconductor, an open-source bioinformatics package using the R statistical programming (Gentleman *et al*. 2004) [13]. More recently Pathak *et al.* 2020 [23] have identified 436 RNA-guidedrecombinase target sites in Bos taurus genome by using dreg program of Emboss package.

Using Blast of NCBI all the identified target sites sequences were mapped against RefSeq Genome Database of Bubalus bubalis (Taxid: 89462) for highly similar sequences to determine whether these RGR target sites are present within any gene or not. Our investigation reveals that 207 RGR target sites lie in genic region and 263 in intergenic region. RGR target sites lying within genic region along with their gene features has been shown in Table 2.Similarly, Pathak *et al.* 2019 [32] has also found RGR target sites to be located both in genic region as well as intergenic region in case of Bos taurus genome. Their findings show that the total number of target sites located in genic and intergenic regions was 173 and 263, respectively (Pathak *et al.* 2019) [32]. Chaikind *et al.* (2016) [6] also searched ENSEMBL (Release 81) to identify which predicted recCas9 target sites fall within annotated genes.

The identified 426 buffalo genomic sites for recCas9 could be significantly expanded by replacing the hyperactivated Beta recombinase catalytic domain with other natural or manmade small serine recombinases that recognize different core sequences. Furthermore, recent work altering Cas9 PAM binding specificity, as well as the recent discovery of numerous Cas9 orthologs, opens up the possibility of increasing the number of potential recCas9 sites (Kleinstiver *et al.* 2015, Esvelt *et al.* 2013, Zetsche *et al.* 2015) [20, 7, 31].

Extending the approach we can use the RGR platform based on hyperactivated Beta recombinase to construct tools that can be capable of specific, seamless genome modification *viz.* integration, deletion and inversion of the identified target sites of the genome map prepared. 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) [21, 11, 13, 10, 14, 26]. Zinc Finger Recombinase with diverse collection of hyperactivated recombinase Gin has been reported for genome modifications (Gaj *et al.* 2013) [11]. And that ZFR architecture is capable to recombine target sequence with extreme specificity and achieved human cells DNA integration into targeted loci (Gaj *et al.* 2013) [11]. Chaikind *et al.* (2016) [6] designed a RGR that operates in mammalian cells, by fusing a dCas9 with hyperactive Gin recombinase and the said RGR proved capable of in situ deletion across two genomic sites. No indels or other DNA modifying products were detected as the RGR catalyzes recombination directly (Chaikind *et al.* 2016) [6]. Standage-Beier *et al.* (2019) [28] generated a RGR named as integrase Cas9 (iCas9) by merging dCas9 into the catalytic domain of Tn3 resolvase derived from transposon Tn3. They have demonstrated iCas9 targets deletion and integration of DNA in both yeast and human cells.

Table 1: Chromosome wise number of RGR target sites in buffalo genome

Chromosome no.	Size (MB)	No. of target sites	Chromosome no.	Size (MB)	No. of target sites
1	202.11	21	13	90.5	11
2	188.95	35	14	83.51	18
3	175.63	48	15	82.16	8
4	165.35	20	16	84.65	7
5	127.68	19	17	73.31	10
6	120.64	25	18	65.93	19
7	117.22	19	19	71.84	10
8	119.77	16	20	68.86	15
9	110.24	20	21	60.88	13
10	104.52	10	22	62.06	10
11	102.29	20	23	51.76	9
12	106.43	15	24	42.45	7
13	90.5	11	x	143.53	21
14	83.51	18	Total		426

Table 2: Locations, Sequences and features of RGR target sites identified in genic region of buffalo genome

Chr no	Start position	End position	RGR target site sequence	Target site feature
1	194249830	194249907	CCTGTGACGACATAGCCTTCTTTGACATCCAAACCAGTTTACCAGGTCAGACTGATTGATACCTTTCTCCTAGGCTGG	Dual specificity tyrosine phosphorylation regulated kinase
1	189243507	189243584	CCAAAAGAAAGACAGGGAGCCAAATGTTCACAGCAGGTTTACCTGTAACAGCCAAGAAAACTAGAAACAACCCAAAGG	Ubiquitin Conjugating Enzyme E2 G2
1	173491460	173491537	CCCCACTGTAACACAGCAGAACCTCACCTTAACGCGGTATACCTGCAAAGATCCCATTTCCAATTAAAGCCATTTAGG	Calsyntenin-2
1	132028651	132028728	CCTATATCCTACCAATGAATAGAGACAGAATAAATGGTATACTATTTGGAAGCAAAGCTTAGGGGAACTATTTGGGGG	Peroxisomal Biogenesis Factor 5 Like
1	108876834	108876911	CCACCACTTGCTAACCAAATGATGTTCAGCAAGGCAGTAAACTTCTCTCAGATCCTCAATTCTCACTGTAAAATGAGG	Rho GTPase-activating protein 31 isoform X2
1	125181	125258	CCAAGAGCAAAATTTTATGTCCCAAGTTTAGAAGTGGTATACTTGGTGTCTTCTTATATTCCAGTGACTAGAATTCGG	Zinc finger protein 385D isoform X3
2	186555754	186555831	CCGCTTTGCTGAAGAGGAATCTGGGGGTTTATAGAGGTTAACTGACAGGTCCAAGGCCTCACAGCAGGAGGTAGATGG	Taste 1 receptor member 2
2	182137224	182137301	CCAGGTACAGTGATTTTTTTTCACCCAGGATTGGTGGTAAACTGAGGCAGGAGGCCTTAGGCTGCTGGTGGCCTTAGG	E2F transcription factor 2
2	178929292	178929369	CCCCATTTGTAAAGGAGGACCCTGCGGTCTGGAGAGGTTAACTACTTACCCAGGGTCACACGCACCAGAAAGTGGAGG	Solute carrier family 9 member A1
2	178692500	178692577	CCTCCTAAGGAGGGGCCTGCTAGATCTGGCCACATAGTAAACCTTTGTGTTGCCTCATTTTGTTCAGTCATGAAATGG	WASP family member 2
2	168534858	168534935	CCCCCATAAAGGAAAGGGCAGATTAGGAGAAAGTAAGTTAACTTAGCCATAAAAGGAGTATTTAGAGATCATTCTAGG	Mitochondrial fission factor
2	162697268	162697345	CCTCCTCAGCACCAGAACTCCGCAAGAATTTAAGCAGTAAACCCACCAGATTCTGCCTGAGTGCACTTTAATGCCAGG	EPH receptor A4
2	149644844	149644921	CCAAGACAGAGCAATATTTAAAATTATAAAGAATCAGTTTACCAAGAGGAAATAATCCTAAATATATATGCGTTTAGG	Parathyroid hormone 2 receptor
2	146742335	146742412	CCAATTTTGCAAATCACATTTTGTGGCTGAAAATTGGTTTACTGCGCCCATCCCATTTTGGAGGCCAGACTTCCCTGG	Par-3 family cell polarity regulator beta
2	129590404	129590481	CCAAAATACTGCAGATTGAGTGGTTTAAAAACCGAAGTTTACTCTCTCACAGTTCTCAAGACTAGATGTCCAAGATGG	Contactin associated protein family member 5
2	125948991	125949068	CCCTCCACTGCTAGCCTCCTGCCCAGCAGACAGCTGGTAAACTGAGCTGCTCCCCACCAGTGAGTAGCAATCCACAGG	Cytoplasmic linker associated protein 1
2	114240718	114240795	CCTACTCCAGGGGATCTTACATCTGTGAAGGGGGAAGTTAACTTATTATATTACATAGGCCTTCTGTTTTTTTTGTGG	Minichromosome maintenance complex component 6
2	105022213	105022290	CCTGGTGTTATTTGAAAAGGAATCCTTCCCAAGAAAGTTTACTGCCAGGGAAGGAAGATCCTCGCAGTAAGGAGTTGG	Glycosyltransferase like domain containing 1
2	90305364	90305441	CCTAGGCTTGACACACACCTTCAGGGTGGAACCCAGGTTTACTGTTTACTTGGCTTAGAAGTCTCACTTTCATTTTGG	Plakophilin 4
2	62280500	62280577	CCTGCACTGTGGCAGGGAGAGGGTGGTAACCTGTTAGTAAACTGTGAAAAGTTAAACAGCAGTAAATGACAGGGCTGG	Integrin subunit alpha V
2	56842340	56842417	CCTTGATTTCCCTCTCAAGCTACTGAGAATCCCTAAGTATACTGAGATTTTTTTTCTGTCTATACATTGCTTTTGGGG	UDP-glucose glycoprotein glucosyltransferase 1
2	25797969	25798046	CCAGATTCAAGCACACAGACCCTTTGGAATTTAAAAGTTTACCCTTGATCTCTACAGGGGCGAGTTTCCCTCAGTAGG	Neurogenic locus notch homolog receptor 4 isoform X3
2	15891284	15891361	CCTGTGTCTTTTATTCCATCCCTCCCATCGCCACCAGTTTACTCTCTCCTCTCCTACCTCTTCCTTGATGCCTGCTGG	Uncharacterized LOC112579558
3	168639618	168639695	CCCATATATTTGAATAAGTGTACTTAATTCAACCCAGTTAACTGTTAAAAAAAAAAAACAGAAGACATTAAGAAGAGG	Astrotactin 2
3	168182482	168182559	CCCCAAAACCACATGCTCGTTTCTCCAAGTCTGGAAGTAAACTTTCTCAAGCCAGGCCCGAAGGATGGGGGTCTGAGG	Pappalysin 1
3	168146774	168146851	CCTGAGGTTTTTGGTTAAAAAAATGTGGTATCATCAGTAAACTGATAACTCTCCACACCCCACCCTGAGCAGCTCTGG	Pappalysin 1
3	166555750	166555827	CCCTGACCCTGGACTGTCTTGTGATACACAGTGTAAGTTAACCACCATCCACTCAGTTGCCCAAGCCAGAAACCTAGG	Transmembrane protein 268
3	165882126	165882203	CCTCCCAGCCTGGACACCTTCGCTGAGCTCCAGTTGGTATACTAGCTTGGGGTAGATGAAGCTGGAAGCCCAGAGGGG	Zinc finger protein 618
3	162011588	162011665	CCCCCGGGGGGCTTAAGAATGCAGTTAAGTTTCCTGGTTTACCCTGAGTTAAGCAGAAATCTTCACCTGGCCAGGTGG	LOW QUALITY PROTEIN: A-kinase anchor protein 2
3	161642114	161642191	CCACTGGTTCCATGTTTTAAGTTACTTCCAGGCTTGGTTTACTTTTGGCCGAATGACCACCCCGTCTAATGATCATGG	Protein tyrosine phosphatase non-receptor type 3
3	161431684	161431761	CCAAGCTTAATCATAAATACAACAAAGCAGATGCCAGTTTACCTGGGAAGGACCACGGTGGGTTAGGACACAGGTGGG	Erythrocyte membrane protein band 4.1 like 4B
3	147415387	147415464	CCAGTGGGTGTTTATTCTTGTTTTTCACATCAGGAAGTAAACTTGCCTTTTATTTCAGGTGGGGATCCCAAACCCTGG	Constitutive coactivator of PPAR-gamma-like protein 1
3	137191457	137191534	CCACACTGTCTTGGCAGTCAGAGGGAAGGCTCCAAAGTTAACCTGTTCCTCTCTCAGGAGACAGGGAGGCCCACCAGG	Epoxide hydrolase 2

3	132673447	132673524	CCTGGTTCCAGGCTAATGCCAGGGAGCCCACACTGAGTAAACTCCTGCTAAGTGCCAAGTGCTGTGTCCACGGGAAGG	Phosphatidylethanolamine binding protein 4
3	109972266	109972343	CCAGCCATCAGTGACCCAAAGTGCACACTATAAGAGGTATACTCTGAGTTCTGAACTGAGCAAACAGTTACAGAGAGG	Transient receptor potential cation channel subfamily M member 3
3	92869118	92869195	CCATGTACCAGATCGGTGAAGGCAGCCGGAATGAAAGTTAACCAGGGGAGGGGAAGTGGTGGTGTTCCAGGAAGATGG	Nuclear factor I B
3	88787519	88787596	CCAATGCAGTATTTTCACAGGTCTTTTTAAGGAAGAGTTTACTTTTGCAAAGAATGAAAGCCACGAAGAGAAGCTTGG	ADAMTS like 1
3	88422306	88422383	CCTTCCTGTTCACTACCATTGTGGGCTGTCCTCAGGGTTAACTGCTGTAACCATTAGCTTGCTCCTCCTGACCAAAGG	Low Quality Adamts like 1
3	70398089	70398166	CCTTCTTCACAGTCCAACTGTTTTCTCTAATTAGAGGTATACTCAGCCATTTTTAAAAATGCTTCAGATTAAAACTGG	Squalene synthase isoform X3
3	49445120	49445197	CCTTCTGGGAAAGCTGGCTTCTTCCCAGTGTGATGAGTTTACTGAGGTTGGTGTGGTGCCTGTGTAGACAGAGCGTGG	Hepatocyte nuclear factor-1 beta (HNF-1β)
3	48349018	48349095	CCACTAGGGCGACCGTCGGTAAGGCTTCGGGACTCGGTTTACTCCCTGATTATGATGACGAGGAAGACGAGGAGGAGG	Zinc finger protein 830
3	45477460	45477537	CCTACAGGGCCCCTGTACTGCCCAGAGGAGAAAGAGGTAAACCCTCCATCCCATCAGACCATGTTAGTAAGGACTGGG	Mitochondrial Rho GTPase 1 isoform X5
3	39579224	39579301	CCAGAGGCAGTATGTTTCAATTACTGTTAAAAATGGGTATACTTCCTGAATGAAGATGCTCCATGGCCTCGGGGCTGG	RAP1 GTPase activating protein 2
3	39025695	39025772	CCTCCTCGTCTTGGCTTGCCCCCACTTTGCAGATAGGTAAACTGACACCCCACACAGGTTGACACAGCGTCGAGCGGG	Tax1 binding protein 3
3	31254696	31254773	CCCCGTGCTTTACCCCACACCTGGACATAGTAGCTGGTTTACTGAGGGGGTATCTGATTCAAACAGTCTCTGCTTGGG	Protoheme IX farnesyltransferase, mitochondrial
3	28978744	28978821	CCATAGAAGTTTGCTAGGAGGGCAGGGACATGGGCAGTTTACCTGGGCTAAAGCTTGGAGGCTGGACTGTGGGGCTGG	Myosin XVA
3	25950061	25950138	CCGAGGTATCAGAGAAATCCCAAGGTAGATAAACTGGTATACTAAGAATGTCATAGCAAGCTCTGTGAGAATGTTTGG	Beta-1,4 N-acetylgalactosaminyltransferase 2 isoform X1
3	24872095	24872172	CCCAGACAGGGTTGGGCTCCTTCTGAGCTGGAACAAGTTAACCATTGACTGGGGTGGAGGTGGAAGGCTGGAAGCGGG	Proline-rich protein 15-like protein isoform X1
3	18253214	18253291	CCTGCATCAGCCAGCTTTGCCAACTCAGGCAAGTTAGTAAACCTTTCTGACCTTGATGTCCTCATCTGTAAAATGGGG	Pleckstrin homology and RUN domain containing M1
3	8229263	8229340	CCCTGCGGGCAGGACGTTGTCTCCGGTGGCCAGAAAGTTAACTCTTCCCTCTTCTGAAGCCATGTGGCCTTTAGAAGG	Progressive rod-cone degeneration protein
4	129286944	129287021	CCTGTGGCTGTTATAGATGGCCTTCCATCTCACTCAGTTTACCTGTAAACACTCTATGTGGATCCCATATGGGATTGG	Ryanodine receptor 2
4	128361803	128361880	CCATTAAGCAAAGAAATAGAAGCTTCTAAAGTAGGAGTTTACTGGGTTGAACTGCAAATGCTCCTGAGATTTTCCTGG	G protein-coupled receptor 137B
4	126489093	126489170	CCTACATGGACACATCATAATGACTCAAAGTCCATAGTTTACTTTCAGATTCCCTCTTGGTACCGTACGTCCTATGGG	Solute carrier family 35 member F3
4	125991190	125991267	CCCATAGAATGCGCAACACCAAGAGTGAACCCCGAGGTAAACTCTGGACGTCTGGTCCTAATGATATGTCAGTGTAGG	Potassium channel subfamily K member 1
4	124110946	124111023	CCAGCATCCCAGAGAGTGTTAGCTGTCACCGCAGGGGTTAACTGGGCTCCTCTAAAGCAGATAGGGTACTCAGATGGG	Disrupted in schizophrenia 1 protein isoform X2
4	114273622	114273699	CCACTTCACTGCTCAGGGCCATGAAAACCTCCCTTGGTTAACCCTTAACCTCATTAAGGCTCTTTGCTAAACAGAAGG	Bardet-Biedl syndrome 10
4	110944916	110944993	CCTAGAGCCACTGTGATGTCTGCAACGATTTTAAGAGTAAACCTGTTTGAGAGTATCTTTGTGGCTCAGACAGTAAGG	PRKC apoptosis WT1 regulator protein
4	63064718	63064795	CCTGTTGCAAGAACTATTACTAGGACAAAGGGGATGGTAAACTGGGGGCTCTGCCAAGCCAGTAACACCCCAGCCAGG	SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 2
4	52232173	52232250	CCTTTTCCTGGGATGTGTTCTACAAAAAATGAGGAAGTAAACTAAGAAAGGGGAAAATAAGAGATCCAGGAAGCAGGG	Host cell factor C2
4	47738765	47738842	CCTGGAACTCAGGAAAGAGCTCTACTTACTATCCTAGTTTACTGTTCAAGGATACAACTTAGGAACATGGTATGGTGG	LARGE xylosyl- and glucuronyltransferase 1
4	22486925	22487002	CCCAAATTATTCCACTTTACAGGTGAGCAAACAATAGTTAACTAATTTTTAAAGAAAGGTTAAGAAATTGTTCCCAGG	LDL receptor related protein 6
5	100789301	100789378	CCCGCTTGTCACCTGGCAGCTGGGGAGGTGGTAATGGTTAACCTGGGAGAGTGCTCCATCTTCCTCTGGTGTCCTTGG	Neuron navigator 2
5	91914839	91914916	CCAAGGAGAAGAACTCACATGACTCTTTTTCCATAGGTTAACTCTCAATGGCCTTTTTTTAAAATTATTTTTTATAGG	Uncharacterized LOC112585287
5	51767190	51767267	CCAGCGGCCTCAGCAGTCCCAGAGTTTGTGCCATGAGTATACCCTGGTGGAGACTTTTTATGTTTGGCACTGTGCCGG	Poly(ADP-ribose) polymerase 1
5	49433811	49433888	CCAGCCCGGGGGAAGAGGCCCCCTGAGCCTCACACAGTAAACCCCTGTGCATTCCCAACCAGGTGGGTGGCATCTGGG	Kinesin family member 26B
5	47500111	47500188	CCAGAGAGGAGGCAACCAACCTGGAGGCAGAAGAAAGTTAACTTTCTTTGTTCCCTTTTTGCTTTCTCTTCGCCCTGG	Serologically defined colon cancer antigen 8 isoform X5

5	43853867	43853944	CCCCTGGAGGACTCATGGAAAGTAGGTGGAGGAACAGTTTACCAAGCCTAAGCAGAAAAGGGGAGCCAGAGGAGAAGG	Selectin P
5	36157315	36157392	CCAATATATAAACACAACTTCAGGGGAAAGGGGAAAGTTTACTGACCACTCATCAATTCAGAAAGCAGGAGGGTGTGG	Arginine-glutamic acid dipeptide repeats protein isoform X3
5	31422453	31422530	CCGAGGTGGACGGGAATGCGGGCGCCGAGGTGAAGGGTTAACTGCGCCTCTGCCAAGCCCATCCTGCAGCGGGGGCGG	PR/SET domain 16
5	15101085	15101162	CCTCAACAGGTCATACGAAAGTGTGCCAAAGGCCTAGTTAACCCTTTTCTCATTCCTGAATTTGCCTCAGTTTCCAGG	Niban apoptosis regulator 1
5	2126668	2126745	CCTTTCGCCTTTCTCCTTCCCCCTTTCTCCAGCAAAGTAAACTCTTCACTTAGCCCCAGAGAGGTGCGGAGAGGTAGG	Nuclear receptor subfamily 5 group A member 2
6	119791576	119791653	CCAGATGGTTGTTTCTGTTGTTTCAAACTTTCAAAAGTAAACTTATGGAGACATCTACCTCCCTGGTGGCTCAGATGG	Uncharacterized LOC112577668
6	113105366	113105443	CCCGGAGCACCGACGTGTCCTACGCCGAGAGCGTGAGTTAACCGCGTCCCCTCCAACACGCAGTTAAGGTTCTTTGGG	Low Quality Protein: transient receptor potential cation channel subfamily M member 8
6	112257239	112257316	CCATTTTCACAGTAGAGAACTCAGAGGTTCAGAGAGGTTAACTGCCTCAGCCAAGGTCACACAGCCTGTAAGCGGTGG	Ephexin-1-like
6	90683672	90683749	CCCATTTTGTAGGTGAGAAATTTGAGGCACACAGAGGTTAACTACTTGTCCAAATTCACAAGCCTTGTTACAGCAGGG	Ubiquitin carboxyl-terminal hydrolase 24 isoform X1
6	19643462	19643539	CCCTGTGAGAATAATCAGGTTATTCAAGGAAAATCAGTTTACCATGCCCTGGGAGATGATAATGGGAAGATAGGATGG	Aryl hydrocarbon receptor nuclear translocator
6	16334349	16334426	CCGTAAGCTTCCATTTATATGAACTATCCAGAATAAGTAAACCTACAGAGACAGAAAAGTGGATCAGCGTTACACGGG	Integrator complex subunit 3
6	15179522	15179599	CCTCCCCAAGACTCAGGTTAGGGGCCGGGCGGGACAGTTAACCTGGAGAGGGGGAGCGGTGCCAACCGGGCGTCCAGG	Ephrin A1
6	14375900	14375977	CCATCACACAGATGAGGAGATTGAGGCTTGGGCGAGGTTAACTGATTTGCTCAAGGCCACGCAGCAGACATGTGACGG	Ucharacterized LOC112585542
6	9212398	9212475	CCCACCTGAAAAGTTGATAAATGTCCCAAGAAGGTGGTATACTTGCTCATTTCTCTTCTGTCCTTTCCAGAACTCTGG	Peroxisomal biogenesis factor 19
6	8142415	8142492	CCTGATGTCACTGTGACGCGGGATAAAGCTACAGGGGTTAACTCTGTGGAGAACAGGACCAGAAACCATCATTGCCGG	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial
6	844435	844512	CCATTAGGAAGATGGCTAAGTCCCATGGATTTCGAAGTAAACTTTTTTCATTCCTAATTAGTCCTAATTATTCCTGGG	LOW QUALITY PROTEIN: adenylatecyclase type 10
6	841421	841498	CCAGGCCATGTCAGTGAAAACCCAGAAACCTAACAAGTATACCTCCAGGGAATTTCCACGTTCTTTTTTCTATTATGG	Adenylatecyclase 10
6	237229	237306	CCAGTCTCCTGTGTCTCCTGCACTGGCAGGCAGATAGTTTACCACTGAACTACCTGGGAAGCCCCCTAATTAGTATGG	T-box transcription factor 19
7	91911815	91911892	CCTCCGGGAAACCAAGGAATGGCTGAGTGCCACTTAGTAAACCCTGGGACCTCTTGTCATGGAGCTTAGAGAGGCAGG	Collagen type XXV alpha 1 chain
7	76503998	76504075	CCAATTAGGAAATCAATTAGGACCTTAGAAATGGTAGTTTACTGTCCTAATTTCAATCAGCCCACAGCTAACATGAGG	Potassium voltage-gated channel interacting protein 4
7	71093609	71093686	CCAGATTAAAGAAGACTGCAAAGTGAGGACACCTGGGTATACTTGCATACTCATTATCAGGACCAAAAGCAAAGCTGG	Stromal interaction molecule 2 isoform X2
7	59195826	59195903	CCCATCATAAAATGACCTCATAGCCAAATATAAAGGGTATACCATCAATTCTAGAAAGCCCAGTTTGTCATTTATTGG	Transmembrane protein 156 isoform X5
7	31003228	31003305	CCATCTGGTTCCAGAAAGAATTTACTGAGGTTATTAGTATACTCCTAGTTTTGGAAGGAACATGAAAACTATTTCTGG	Solute carrier family 4 member 4
7	29958280	29958357	CCCTTTTCACATATGATGAAACTGAGACTTTGGGAAGTTAACTGTTTCTTAGAGGTGGACTCAGTGCTCAGACTTGGG	ADAM metallopeptidase with thrombospondin type 1 motif 3
7	26325500	26325577	CCTACATAGACACATCACAATCACCCCAAGTTCATGGTTTACCTTGGGGTTTATTCTTGGTGTGGTACTATTCTATGG	Coiled-coil domain-containing protein 158 isoform X7
7	15317672	15317749	CCCCTAAAGTCTGTCCTCTGTACATTTTTTTAACAAGTTTACTTGGCTGTGTCGTGTCTTCATAGCAGCACGGGGGGG	Hydroxysteroid 17-beta dehydrogenase 11
7	227721	227798	CCGGGGCCCACATCTCCGCCTGTCTGGGCTCTGAGAGTTAACCTCCTGCTGGACTTGGAGGTCTCGCTGCGGCCTGGG	Polycomb group ring finger 3
8	116433055	116433132	CCCCACAAGTTCCAGAGTTTTGTTTTTAATCTGGCAGTTTACTGGTAAATGTGCAGTGAGTCCAGGCTGTCTATTTGG	Dipeptidylaminopeptidase-like protein 6 isoform X2
8	106169221	106169298	CCCGAGTTATTGAGCTGCAGCTTCTTAGAACTCAGGGTAAACTGGCCGTCTTGGGGCCTGGAAGCTTTGAAGTTCTGG	T cell receptor beta variable 6-1-like
8	103023694	103023771	CCTCAGCTCAATAGGAACACAGAAGTGTTCACAGAAGTAAACTTCTTGCTTGAAGGGAGCTGGGAAAAAATTCTTTGG	Thromboxane A synthase 1
8	90888347	90888424	CCACTTGCAGAGTACTTACTCTATTGCAAAAGCAAGGTAAACTGTCCTTCTGTGAGGAGGGAGATCTTTACTTGGAGG	Glutamate metabotropic receptor 8
8	67987412	67987489	CCCCCCAGTTATTTTTCTAAAACATGGAACATATCAGTTTACTCCTTCTAAGGCTCTCTATTGGAGAACATTTGGAGG	Juxtaposed with another zinc finger protein 1 isoform X3
8	66715809	66715886	CCTCCCCAAACCTTACTTCTCTCAACTGTGAAATGGGTAAACTCCTACTATGAAGGCGCTCGTTAGCACTAAAGAAGG	Chimerin 2

8	63981841	63981918	CCTTTCCACTGTGGCCATTTCTGCCCTGGAGGACTGGTTTACCACTCCTTATGTGCCCAGGCAAAATTGTTTTTTAGG	Phosphodiesterase 1C
8	59733675	59733752	CCCAAGAGCAAGGGCTAATTTTGACCAGTGAAAGTGGTTAACCCTTTGGCACATCCTCTTCTGCTGCTGTTTATGAGG	Engulfment and cell motility 1
9	99337528	99337605	CCGTGACTATAGTTTCCTGCACCTCCTGAAGCCTCAGTTTACCCTCTCTGTGAAATGGGCATAAATGTTTGGAGGAGG	Coiled-coil and C2 domain containing 1A
9	96447088	96447165	CCGGGCATCCTTGTTGAACAGAGCCTGGGCTTGGGAGTTAACCAGGTCAAAGTCTGAGCCTCCATCCCCCTATTTAGG	Ribonucleoprotein, PTB binding 1
9	91435899	91435976	CCCATTGGAGACTTGACAAAACTGAGGCTCAGAGAGGTTTACTCCCTTACCCAAGGTCACGCAACTCCACCCAGTTGG	LOW QUALITY PROTEIN: cyclic AMP-responsive element-binding protein 3-like protein 3
9	90293683	90293760	CCTGCCCGAGTCCACCACGTGCAGCTCACTTCCCTGGTAAACCTGGGACCATGTGTCTCCAAAGACCTGGGCCATGGG	G protein subunit gamma 7
9	66035753	66035830	CCTGGACTGTCTGCGGCTAGGCAGCTTGGAAGCCCAGTTTACCCTCCCTGGCTCATTCCAGCTGACAGGCTGTAAAGG	Follistatin like 4
9	65576887	65576964	CCTCTTCTTCATTGGGTTTTTGGGAAGAATGAGTTGGTTAACTTCCACTGTGCGGAAGGGTGCATGATATGGAGAAGG	Voltage dependent anion channel 1
9	57843021	57843098	CCTGGAATAATGCCCATGGCTAACAATAAAGAAATAGTTAACTAAGCTATGATGTATCTGATCCATCCATTCAACTGG	Uncharacterized LOC102395002
9	46317865	46317942	CCTGGTATCGTGTGGAGCTCAATTTCAAACCTGAAAGTAAACTCAGTGAGCCTGACTACTCATGCATCAGAGTGCTGG	Glutamate receptor 1 isoform X2
9	29688873	29688950	CCTGGTCGGTCTGTCAATTTGAACATTTAGTGCCAAGTAAACTGCCTGCTAGAACAAGAAATGAACACTCTTCACAGG	Ras protein specific guanine nucleotide releasing factor 2
9	15747113	15747190	CCTGAAAATCTAATTCCAGATTGACTTCACAAGAAAGTTTACTGGTTTATGTATTTGAGAAACCCAGCAATAGCATGG	Arylsulfatase family member K
10	60662062	60662139	CCCAATCCCTCCCAGCATCAGAGTCTTTTCCAATGAGTAAACTCTTTGGACACCCAGGACTGATTTCCTTTAGAATGG	Autophagy related 5
10	51514780	51514857	CCAGGGGTCTGAAACATCCTGGGTGCTAAATTGATGGTTTACTGGTATTAATGTTAGTTTTTATTTTCCAAGCAGAGG	Kelch like family member 32
10	27008958	27009035	CCTCACAAGTTTACTAACCAGGGGTCATGCACAGAGGTTTACCACCAAGTCAACTGGAAGGAAGAAGATGGCAAGTGG	Uncharacterized LOC112587399
10	1755474	1755551	CCGAGGAGGGGGAGGGGAGGGCGGCACCTCACCTGAGTTTACCCGGGGCAGGTACGCGGGGACGGCGGGGCTGGCAGG	Afadin, adherens junction formation factor
11	100462745	100462822	CCTGAGCAAGACATGACTATGGATGAGCTAGCCTTGGTTAACTCGCTGAGCCGCTCTGAACTGTCCCTCCGTCTCTGG	Uncharacterized LOC102399489
11	81432102	81432179	CCCTGAAGTCTGGAAAAACAATGAGCTTAAAACAAGGTAAACTCTCCCACTCTTTATCCACCTTTCCCATTCTTTAGG	Zinc finger homeobox 2
11	80879037	80879114	CCTGGCTGTTTGCTTCACACAGTTCACTTCCCATAAGTTTACCCAGAAAAGAATAGGGGAAAAGAGAGGCCATCTGGG	Low-density lipoprotein receptor-related protein 10 isoform X2
11	74449726	74449803	CCTCAGTAAAGTCGTAAAATTCTTTAGAACTAGCAAGTAAACTGTTCATGACTAGATAGTCTTGCCCTACTCACTTGG	Ryanodine receptor 3
11	65105287	65105364	CCACCTCCTGAATTCTGCCCTGCTTTCAGCAAAGCAGTTAACCTTCTCAATTGTGATATCATGATGTACATTCTCAGG	E3 ubiquitin-protein ligase UBR1 isoform X2
11	57991216	57991293	CCATTAGTCCTCATCGGACAAGTTCTGGTCCAGGGAGTAAACCATGAACATCACATCTGGTCGAGAGGGGACACAGGG	Ornithine decarboxylase antizyme 2
11	57694972	57695049	CCGTATGATCCCATTTATATGAAATACCCAACCTAGGTAAACCATGGCGGTAGGAGGCCGATGAGTGTTTGCCAGCGG	Casein kinase 1 gamma 1
11	47183442	47183519	CCCGTGACTCAAGTGAGCCCACAGCACCCTGAGTGAGTAAACTTAGTTCCCTTGGGACTGGCCTCCTTCACTGGCGGG	LOW QUALITY PROTEIN: protein unc-13 homolog C
11	31313854	31313931	CCCTAGGTGAAGAACTGTTAATCTTGTTAAATCAAGGTTAACCCAGAACCAGGAGCTGAGAGTTGACCATACCTGGGG	Reticulon 1
11	22754549	22754626	CCAGGATTTTTGTGGTTTCTTCTGCCTGGGACCAAAGTAAACCCTCTTGGCTGGTGGAATGGGAGAAAGAGGAGGAGG	ZFP36 ring finger protein like 1
11	15918164	15918241	CCTGACTCTCACCTGAGCTCAGGGGAAGGGAAGTGGGTAAACCTGCCTAGTGACCCCAGCCTGGAAGCTGATCCGGGG	Intraflagellar transport protein 43 homolog isoform X2
11	318417	318494	CCTGATATGACTGGCAACTGGAGATAAGCTGAACAGGTAAACTGATTTTAATAAAGGGCTTAAGACAGGCTATCAGGG	Eukaryotic translation initiation factor 3 subunit J
12	101928343	101928420	CCGCCGTTCTCAACCCTGCTCGGCCTCGCGCCGGGGGTTAACCTGCCTCCATGCCTCGGCTCCACAGTCGGCCGAGGG	Senataxin
12	101777180	101777257	CCATGACCCCAGGGGTCAGCGCCCAGGCCCCAGAGAGTTTACCCCCAAAGTTCTAGAAGCATCATCCAAGGGGGAGGG	Netrin G2
12	98163462	98163539	CCCAAGAGGACAGCAGGCTGTTGGAAGGGATCCAAAGTAAACTTGCTGGCTTTGACCCAGGAATGGTCCAAACCCAGG	Family with sequence similarity 102 member A
12	71978874	71978951	CCTGCACCAACACATCACAATCACTCAAAGTCCATAGTTTACCTTTGGGTTCACTCTTACTGTGGTACATTTTATGGG	General transcription factor 3C polypeptide 2
12	70171452	70171529	CCGGAGAAGAACTGTATGAACTGGGCTTTGCCAGGGGTAAACTCAGACAGGGAGACATGGAGGGTGACCCAGGAGAGG	ALK receptor tyrosine kinase
12	38676614	38676691	CCTACAGTGATAGATCCTAATCACTCAAAATCCACAGTTTA	Coiled-coil domain containing

			CCTTGGGGTCCATTCTTGGTGTTGCACATTCTATGGG	85A
12	27778957	27779034	CCTGTTCCTTTGACTAATAGTTTTATGGTGCCCAGAGTTAACCATTAAAGTCAAGCTATTTACATGCGCAAGACATGG	Protein kinase C epsilon
12	10546916	10546993	CCCAGCCTGAACTTGTTCAAGCCAACAGAGAAACGAGTTAACCCTGGGGGAAGGTGCAGCAAAGGGAGTCTGTCTGGG	Tetmethylcytosinedioxygenase 3
12	2935495	2935572	CCTGGTATCAGAGGGTAGCCTCTGCCCACCACCAAAGTTTACTGGGCACCTACTCTGTGCCTGGTTGGCACAGGCAGG	Tyrosine-protein kinase ZAP-70
13	67499132	67499209	CCCTCCCTTATCAGTTGGGTAACTTTGAGCTGGCTAGTTAACCTCCAGAGACTAGAAGTGCCATCTGTAAAATGAGGG	Proline and serine rich 1
13	63027374	63027451	CCATGACAGGGAGCATCATAACTTGTCATGTGCTGGGTTTACTCTGCTTGGGGGAGTAAGGGAAAGTTCAACAGAGGG	StAR related lipid transfer domain containing 13
13	18160059	18160136	CCTAAAGCCCTAAACTGGGCCAAAACCTTAATAGTAGTAAACTTGGTTCTCTCCTCCCAACTACATGACACTACCAGG	Multidrug resistance-associated protein 4-like
13	17182573	17182650	CCAGGTAGTGTCCTGTAGTTGGATACAGAGAACCAAGTTTACTATTAAGCTTTCTGTACAGCTTAGAAAAGGCTGAGG	Multidrug resistance-associated protein 4-like
13	11564561	11564638	CCAGGGTATCAGGAAGCACAGTTGGACTTGGTGACAGTTAACTGGTCGTTTATCTTTTGAGAAGACAAAACTATAGGG	Serine/threonine kinase 24
13	699743	699820	CCCGATGACGTCTGCGGGGGGCGGGGGGGGGGACAGTAAACCCGGGGCTGAGATCAGGGAGCAGAGGGGCCCCGTGG	MCF.2 cell line derived transforming sequence like
13	396268	396345	CCGGCCACGAGGGACCCAGCAGAGGGCACGTGGCGAGTTAACCCCCCTTTCAGGACTCTCACAAGACTGGCGCCTCGG	Transcription factor Dp-1
14	66024198	66024275	CCAAAAAATACTTTTTAAATCATCTCTTACGTGCCAGTTAACTGTGTACTGAGGATAAAAAGATGAACAACTCCTGGG	Ablinteractor 1
14	60130230	60130307	CCCCTGCCCCACCACCATGATCTTTTTAAACCCTGAGTTAACTGATTTCATGGCTTTATCACCCAGAAGCTGCAATGG	Phosphatidylinositol 5-phosphate 4-kinase type-2 alpha
14	58668568	58668645	CCCACTGCCCTGCAAGTTTCAACCAGAAAGAAGGAAGTTTACCAAGAGAAGCATTTCCCAGCCTGGAGAACCCCAAGG	KIAA1217 ortholog
14	55218348	55218425	CCAGGGCGTGCAGGCTTCAGTAGTTGTGGTACACAGGTTTACCTGCCCCGAGGCATGTGGAATATTCCTGGACCAGGG	FERM domain containing 4A
14	44641839	44641916	CCTGGAAACCTGGGTTCCCTGGATCTTGACTGAACAGTAAACTTTCCAGGTGGTTCTGAGAAGACCCTAAGATGGAGG	Sodium/potassium/calcium exchanger 3 isoform X1
14	37483266	37483343	CCTAACAAATACAGTACAAATACAGTAGTGGCCCCAGTTTACTGTGTCCCAGTCCTCACTGTGTTGATTGTGTAGTGG	WD repeat domain 37
14	36628556	36628633	CCATATGTCCTTCCCCAGCCCCCTTCCTGAGTCTAAGTTTACTCATCCTGTATCCATTTCTGTAGAGCCTGTTTTGGG	Ras association domain family member 2
14	24915449	24915526	CCCAAAGACTTGGGGGCTCACATGGGTCAGGACTAAGTTAACTCAGCCTAGGCTGAGAAGGCGCTGAGGCTAAGCAGG	CCCTC-binding factor like
14	17302101	17302178	CCTAACCCTGCCCCTATTGTGTAACCCAGGTCATGGGTATACCACTCTGAGCTTCAGTTTTCACATCTATAAAATGGG	SRC proto-oncogene, non-receptor tyrosine kinase
15	22953124	22953201	CCATCATTCCTACATTGTGGTGTGTCAGTAAACCCGGTTAACCCACATTTGTTCAACAATAGCTATGTTACAGATGGG	Zinc finger protein, FOG family member 2
16	70880112	70880189	CCTGCAGACATCTCCACAACTTCTGGGCTCTAAAAAGTTAACCACTGTAGTGGTGCAAAGTATGTGTGTCTGGGGAGG	Mastermind-like protein 2 isoform X2
16	39711205	39711282	CCTCATATGTCATGATTACTTACATTTCAAAGAAAAGTTTACTTTTAATTTTTTTGGCCATACCGTGCAGTTTGCAGG	PPFIA binding protein 2
16	18139493	18139570	CCAACGAGACTATGACAGCCATGAGGGGAGGGCTGGGTTTACTGGTCTACCTTCTATACTCACTGCCCCAGACAAAGG	Low density lipoprotein receptor class A domain containing 3
16	7648674	7648751	CCTGCTTGGCACCTCACCCAAAGCCTCTCTGGGTCAGTATACCAGCTGGCAGGAGTGGGGTGCCTTGGGAGAGGGTGG	ADP-ribosylation factor GTPase-activating protein 2 isoform X1
17	57451145	57451222	CCATCTCATACCAGTCATAATGGCTGCCATCCAAAAGTTTACTAACAATAAATGCTGGAGAGGGTGTGGAGAAAAGGG	Inositol polyphosphate-4-phosphatase type II B
17	21586363	21586440	CCAGAGAACGGAGTTCCAATTCCTGGATGGAAAAGGGTTTACTTTGAGTCTTTTCCCGGTGCAGGTCCGCAGCTGAGG	Nuclear receptor corepressor 2 isoform X3
17	20656816	20656893	CCCTATTCTCCACTAGACAGATGAGCTGGCATCCAAGTTAACTCACATATTTACATATTTTTTAAGAGAGAAAGGTGG	M-phase phosphoprotein 9 isoform X3
17	9008477	9008554	CCCTGGCACATTGTAAATGTTAGAGAGTTGAAAAAAGTTAACTAGTCAAAGGAGGGGGGTGCAGATATTCCAGAGAGG	Ubiquitin carboxyl-terminal hydrolase 30 isoform X1
17	120627	120704	CCCAGCGCCAGGGCAGTGGTGAGGGCTGAGGAGGGGGTATACCTATGGTGCTGACTGCAGCACCCCCGCCCCCCAGGG	Coiled-coil domain-containing protein 188 isoform X3
18	56735494	56735571	CCGCGGAGTCCGAAGTCCGGAAGGGCCATGGAGTGGGTAAACCATCCCCACGCCCAGGGGATCGGTGGGCGAGGTTGG	Synaptotagmin 3
18	51961202	51961279	CCCACATTTTGAATGAAGTAACTGAGATCCAGAGAGGTTAACCGGCTTGCCCAAGGTCACACAGCCATTAAGTGGTGG	SMG9 nonsense mediated mRNA decay factor
18	49817513	49817590	CCAGCTGTACCTAGCAGTGGGATCTTGGGCAAGTCAGTTAACCTTTCAAGACCTCCGTTTTCTCATCTGGAAAATGGG	RAB4B, member RAS oncogene family
18	35619157	35619234	CCGCTGAGTTTCATGATGGTTTACTATACAGTAACAGTTAACTAATACAGCCCTTTAAAGAGGAAATGAAGTGGGAGG	Cadherin 3
18	16321038	16321115	CCACTGTTTTACCCCAAGATAAGGGCATAGAAGAAGGTTAACTGTGCAGTTCCAGCAGGAAGGATTTGTGGGACAGGG	LOW QUALITY PROTEIN: lon protease homolog 2, peroxisomal

18	11272429	11272506	CCTTGGCAGCATGGGGCTTCCCTTAAAAAATAGTAGGTAAACTTTGGCAAAAACAGCAGCCACGTAGTCTTCTGAAGG	Gse1 coiled-coil protein
18	6801880	6801957	CCTCGCACCTGTTTTAAGTTACATGGTCTTGGGCAAGTAAACTAGGTTAGTGCTAGATCCTTAGGGCTTCCCCTGTGG	Uncharacterized LOC112580196
18	1859064	1859141	CCAGGCATGCGATGAGTGCTCAGCAGAGGTTACTGGGTTTACTCCGAGTCAGGCTTCTGCTGATGGCCCTTTGCCCGG	Fatty acid 2-hydroxylase
19	70812642	70812719	CCCCATTTCTGAGCCCCTGCACTGGTGTCACAGAGAGTTTACTTCCTGGACGGCAGCCTGGTTTCCCGAACACCCTGG	Lysophosphatidylcholine acyltransferase 1
19	61196547	61196624	CCTGACAGGTCAGGAGATTCTCCTATCAGGTAATTAGTAAACTTGCAGGTAGCTGGCAAGATCTTCCTTTCTGAGGGG	Catenin delta 2
19	31324188	31324265	CCACCCTGCTCACCGTCGACAATCAGGGTGATGTCAGTAAACCGGTCCTGCTCTCGCTGTTCATTCAATCGGTCCAGG	Zinc finger protein 131
19	3729711	3729788	CCTCAAGCATCTACCCCCATAACCAGAGCCTGCCTAGTTTACTACTTTGTGCTCTCACCAACACCTCTGACTTTACGG	Serine/threonine kinase 10
20	62204379	62204456	CCATTTCCTTTCATTATATACTCTCAGAGGTTAAGGGTATACTAATAATGTCATTGACTGCTTCCTAATAACAAGTGG	Leucine-rich repeat-containing protein 28
20	53223260	53223337	CCGACCTCCAGCGAAAGGCAAGGCACAGAGCTGAGGGTTTACTTCCCAGCTAAGGCTGGAGATGCCACGTGGGCGGGG	A-kinase anchor protein 13
20	50403192	50403269	CCTCTTGATGGCTTTGTGAAAGTCTTTTGTTCCTGGGTTAACCCCTTTTGTCTCATGGGACTCGCCTTGGTCTTGAGG	Neurotrophic receptor tyrosine kinase 3
20	47588107	47588184	CCTTGTTTATTGAACCAGTTCCCTTTTAATAGATAAGTAAACTTGCCTGTTTTGAGACCACTTTTGAAGGAACTCTGG	BLM RecQ like helicase
20	33583272	33583349	CCTGACCCATTTCAGTATTACTGTACCTGACTGATAGTAAACTATTGCTGATCAAACTTTATGAACTTAACCTAATGG	Uncharacterized LOC112580847
20	4957010	4957087	CCCCAGTTTAGCTGCTGTGTTATCTTGGGCTGGCCGGTTAACCCCTCTGTGCCTCTCTCCTCATGGGTCAAGCGGGGG	Coiled-coil domain containing 85C
20	4572055	4572132	CCTCCAGTCATCAAAATACTAGCAAACTGTCAGAAGGTAAACTTTTCTGAAGGAAACAATGTGGACCAATAGGAGAGG	Enah/Vasp-like
21	56552856	56552933	CCCAAGTGTGAAAACAATACGCTGTGATTAGGATTGGTTTACCCAGGGTCGTGGTAGTTTGGACTCGTTTTGCCTGGG	Raf-1 proto-oncogene, serine/threonine kinase
21	43149058	43149135	CCTTTAGGAATATTATTATTCCCATTTTAGAGAAGAGTAAACTGACTTTCAGAAAGACATTTAGTAACTTTCCCAAGG	Filamin B
21	28370731	28370808	CCTTTGACTATTTCGTTTCTGCATGTTCTTCTCTGAGTAAACTAATTCTTCGGCTGAAGTTTTTCCACAAACAGAAGG	LOW QUALITY PROTEIN: E3 ubiquitin-protein ligase PDZRN3
21	10214522	10214599	CCACTGAGTTCATGGTAATTTGTTAGGACAGCAGTAGTAAACCAATACATATCTCAAGTTGCAAAACTCCAATTGAGG	SH3 and cysteine rich domain
22	29307342	29307419	CCAATGCCTGGTTTATCACCAAGACTGTAGTTAACAGTAAACCTCTGCGTTTCACTGTGGTGGTCTCAGCCTCTGGGG	Laminin subunit alpha 3
22	29065116	29065193	CCAGGGGCCTCAGGAAGACTTGAGCGGAGATACCGAGTTTACCCTCTTTGATTTTTTTGGCCACACTGAGCAGCTTGG	Regulator of MON1-CCZ1
22	18897287	18897364	CCACAAACTCTACTTCCTTCTCCACACCATGGACAAGTAAACTGCTAAGCCCAGGCAGTAAGCACAGCCCCAGGGAGG	Nucleoporin SEH1
22	3898070	3898147	CCTTCCAATTGACTGAGCCTGGGAAAGAGCCAGTGAGTATACTTATTTTTCTCAGAAAGCTTATTTGTAATTTACTGG	Protein ERGIC-53
23	46254355	46254432	CCAGGCAACTGGGAGTGCAGGTGCGGTCAGCTGAAGGTATACTCGGATCTCCGAGGTCACGCTGTCCTCTGATAGCGG	Dedicator of cytokinesis 1
23	40286805	40286882	CCTGGTGACCCATATCCCTGCATTCCTGATTTGTCGGTAAACTGACACTGATCACTGCAGTCACATGATAGCCCAGGG	WD repeat domain 11
23	22541083	22541160	CCACTACTTACTAACTGTGTGAGCCTGGACAAATCAGTTAACTTTTCAGTGCTTCATTTTTCTCATCTATACAATGGG	UPF0668 protein C10orf76 homolog isoform X3
23	17557615	17557692	CCCTGATCCAGGAAGATGCCATGTGCCTTGGAGCAGGTAAACCCATGTGCCTGTTGAACCTGTGCTCTAGAGCTAGGG	Tolloid like 2
23	8544543	8544620	CCTGATCTCCAGGGAGGGAAAAGAGCTAGAGAGTGAGTTAACCACCCATGGCCAGTGATTTAATCAATCATGGCCAGG	Putative neutral ceramidase C isoform X1
23	8153168	8153245	CCTGTTTCAACCTACAGTTTTATTTTGAGCAAGTTGGTTAACTTCTCTGTGCCTCACTTTCTCTATCTGAAGAAATGG	cGMP-dependent protein kinase 1 isoform X2
24	20921873	20921950	CCTATAGATACTAAAAAGTCCTACCAGGTGCAAGAAGTTAACTACCCAGTGACCAGACTTTAGCCATGACATACGTGG	Protein kinase C beta type isoform X1
24	19187581	19187658	CCTATGACTGTCCTGTTACTGGAGCAGATCATAATAGTTAACTGTCTGGAATGCTAAGGTGATAACTGGGTATTGTGG	Heparan sulfate-glucosamine 3-sulfotransferase 4
24	15355962	15356039	CCAAGATGTGCCAGTTATCGGGCGGCGCTGCCCCCAGTTTACCACGGCGGAGCCAGAGCAGCCTCCCTGGATGCCTGG	Lysine acetyltransferase 8
24	13422197	13422274	CCCAATCCCTCCCAGCATCAGAGTCTTTTCCAATGAGTTAACTCTTCGCATGAGGTGGCCAACCCTTGGACTGCAAGG	Polypeptide N-acetylgalactosaminyltransferase 17
24	4392879	4392956	CCTGGACTCCCAGGGCCTGCCTGCGTCTGCCTGTGGGTTTACCAAGGTTGGAGAACTGATTTGCCTCCTGGTTTGTGG	Cytohesin-3 isoform X1

X	131437453	131437530	CCCTCTCCTTTTCTTCCGCCACCTAATTCACTCCAGGTATACCTTCTGTGAATTTTATGGAACAAACACAGTGTCTGG	Chromobox protein homolog 3 pseudogene
X	122910380	122910457	CCTTTTGATTCATGGATAATCATAGAGATTGGCTCAGTTTACTGTGTGGAAATTCTTAGTGTGGTAGTGGTATTAAGG	Teneurintransmembrane protein 1
X	85312670	85312747	CCAAATGCCGCTAGCACCAAGATGATGTTTTTAAAAGTTAACTACTTTCAACAACACAATTGTAGTTTTATCCAAGGG	Collagen type IV alpha 5 chain
X	81764248	81764325	CCCCCATTTCACAAATGAGGAAATAGGATCAGGCAAGTTAACTTGTATGCCTGATATTGTACAGCTAGTAAATGGTGG	Transient receptor potential cation channel subfamily C member 5
X	48103984	48104061	CCTGGGTTCAAGATACAAGAAAGCTGTATTCAGAGAGTATACTGATGGCACATTCAGGGTCCAACGGCCAAGGACTGG	Hephaestin isoform X1
X	31018033	31018110	CCTGAACCGAACCCATGCTTTTACAGGGACTGGGCAGTTTACTTCGAATGATGATTAAATTTATCATCCAAACCTGGG	Dystrophin isoform X3
X	3418561	3418638	CCCACAAGAGTGACCTGGAGCAGCGATCAAGCCCAAGTTAACTTTGTATCTCCAAGGCTATGTTCCTCAGATGAGGGG	Neuroligin-4, X-linked

Conclusion

Implementation of RGR platform based on hyperactivated recombinase Betaas a tool for genome modificationrequires a detailed map of RGR target sites throughout the buffalo genome. The present map envisages the complete information regarding location of RGR target sites in buffalo genome and these RGR target sites may prove especially useful for targeted manipulation of DNA *viz*. integration, deletion and inversion to bring genetic improvement in buffalo genome.

Acknowledgment

The authors are thankful to Director, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, (U. P.); India for providing all necessary facilities to carry ought the research work.

References

1. Akopian A, He J, Boocock MR, Stark WM. Chimeric recombinases with designed DNA sequence recognition. P Natl Acad Sci USA. 2003;100(15):8688-8691.
2. Alonso JC, Weise F, Rojo F. The Bacillus subtilis Histone-like Protein Hbsu Is Required for DNA Resolution and DNA Inversion Mediated by the β Recombinase of Plasmid pSM19035. J Biol Chem. 1995;270(7):2938-2945.
3. Brookhouser N, Raman S, Potts C, Brafman D. May I cut in? Gene editing approaches in human induced pluripotent stem cells. Cells. 2017;6(1):5.
4. Buchholz F, Stewart AF. Alteration of Crerecombinase site specificity by substrate-linked protein evolution. Nat Biotechnol. 2001;19(11):1047-1052.
5. Canosa I, Rojo F, Alonso JC. Site-specific recombination by the β protein from the streptococcal plasmid pSM19035: Minimal recombination sequences and crossing over site. Nucleic Acids Res. 1996;24(14):2712-2717.
6. Chaikind B, Bessen LJ, Thompson DB, Hu JH, Liu RD. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. Nucleic Acids Res. 2016;44:9758-9770.
7. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, *et al*. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods. 2013;10(11):1116-1121.
8. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013;31:822-826.
9. Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Methods. 2011;29(9):816-823.
10. Gaj T, Mercer AC, Gersbach CA, Gordley RM, Barbas CF. Structure-guided reprogramming of serine recombinase DNA sequence specificity. P Natl Acad Sci USA. 2011;108:498-503.
11. Gaj T, Mercer AC, Sirk SJ, Smith HL, Barbas CF. A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic Acids Res. 2013;41:3937-3946.
12. Gaj T, Sirk SJ, Barbas CF. Expanding the scope of site-specific recombinases for genetic and metabolic engineering. Biotechnol Bioeng. 2014;111:1-15.
13. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):1-16.
14. Gersbach CA, Gaj T, Gordley RM, Mercer AC, Barbas III CF. Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucleic Acids Res. 2011;39:7868-7878.
15. Grainge I, Jayaram M. The integrase family of recombinases: Organization and function of the active site. Mol Microbiol. 1999;33(3):449-456.
16. Grindley ND, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. Annu Rev Biochem. 2006;75:567-605.
17. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. Nat Med. 2018;24(7):927-930.
18. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, *et al*. p 53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. Nat Med. 2018;24(7):939-946.
19. Karpinski J, Hauber I, Chemnitz J, Schäfer C, Paszkowski-Rogacz M. Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. Nat Biotechnol. 2016;34(4):401-409.
20. Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 2015;33(12):1293-1298.
21. Mercer AC, Gaj T, Fuller RP, Barbas CF. Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic Acids Research. 2012;40:11163-11172.
22. Orthwein A, Noordermeer SM, Wilson MD, Landry S, Enchev RI. A mechanism for the suppression of homologous recombination in G1 cells. Nature.

2015;528(7582):422-426.
23. Pathak SK, Sonwane A, Kumar S. Exploration of RNA-Guided Recombinase Target Sites for Hyperactivated Recombinase Beta in Bovine Genome. Indian Journal of Animal Research. 2020;54(9):1086-1097.
24. Rojo F, Alonso JC. A novel site-specific recombinase encoded by the Streptococcus pyogenes plasmid pSM19035. Journal of Molecular Biology. 1994;238(2):159-172.
25. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology. 2014;32(4):347-355.
26. Sirk SJ, Gaj T, Jonsson A, Mercer AC, Barbas CF. Expanding the zinc-finger recombinase repertoire: Directed evolution and mutational analysis of serine recombinase specificity determinants. Nucleic Acids Research. 2014;42:4755-4766.
27. Smith MC, Thorpe HM. Diversity in the serine recombinases. Molecular Microbiology. 2002;44(2):299-307.
28. Standage-Beier K, Brookhouser N, Balachandran P, Zhang Q, Brafman DA, Wang X. RNA-guided recombinase-Cas9 fusion targets genomic DNA deletion and integration. The CRISPR journal. 2019;2(4):209-222.
29. Thyagarajan B, Olivares EC, Hollis RP, Ginsburg DS, Calos MP. Site-specific genomic integration in mammalian cells mediated by phage φC31 integrase. Molecular and Cellular Biology. 2001;21(12):3926-3934.
30. Turan S, Zehe C, Kuehle J, Qiao JH, Bode J. Recombinase-mediated cassette exchange (RMCE) – A rapidly-expanding toolbox for targeted genomic modifications. Gene. 2013;515:1-27.
31. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759-771.
32. Pathak D, Gandhi D, Gupta A. Self-supervised exploration via disagreement. In International conference on machine learning; c2019. p. 5062-5071. PMLR.