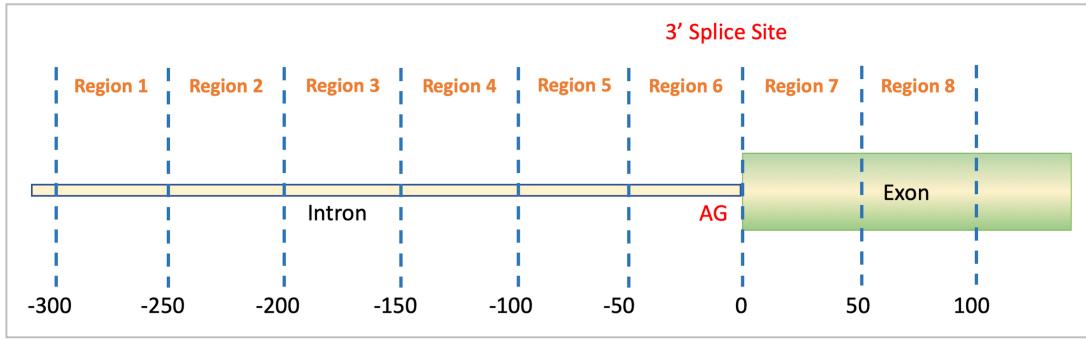


## Motif location analyses



Eight regions are defined for the analyzed sequences. Region 1, Region 2, Region 3, Region 4, Region 5, and Region 6 are located in the intronic region at the upstream side of the acceptor dinucleotide, i.e., AG., Region 7 and Region 8 are located in the exonic region at the downstream side of the acceptor.

	<b>id</b>	<b>motif_id</b>	<b>start</b>	<b>real_start</b>	<b>stop</b>	<b>real_stop</b>	<b>strand</b>	<b>p-value</b>
1	1	MA0528.1	356	52361541	376	52361561	+	0.0000000000000595000
2	2	MA0528.1	317	10642286	337	10642306	-	0.0000000000002870000
3	3	MA0528.1	9	201874860	29	201874880	+	0.0000000000002870000
4	4	MA0528.1	329	201874860	349	201874880	+	0.0000000000002870000
5	5	MA0528.1	335	1641766	355	1641786	+	0.0000000000002870000
6	6	MA0528.1	335	1708898	355	1708918	+	0.0000000000002870000
7	7	MA0528.1	359	52361538	379	52361558	+	0.000000000004020000
8	8	MA0528.1	332	1708901	352	1708921	+	0.000000000007890000
9	9	MA0528.1	332	1641769	352	1641789	+	0.000000000007890000
10	10	MA0528.1	233	20342364	253	20342384	-	0.000000000008290000

```

1 -- Region 1
2 SELECT motif_id, count(*) FROM motif_locations WHERE start <= 50 - (char_length(motif_id) / 2)
3 GROUP BY motif_id;
4 -- Region 2
5 SELECT motif_id, count(*) FROM motif_locations WHERE start > 50 - (char_length(motif_id) / 2) AND start <= 100 - (char_length(motif_id) / 2)
6 GROUP BY motif_id;
7 -- Region 3
8 SELECT motif_id, count(*) FROM motif_locations WHERE start > 100 - (char_length(motif_id) / 2) AND start <= 150 - (char_length(motif_id) / 2)
9 GROUP BY motif_id;
10 -- Region 4
11 SELECT motif_id, count(*) FROM motif_locations WHERE start > 150 - (char_length(motif_id) / 2) AND start <= 200 - (char_length(motif_id) / 2)
12 GROUP BY motif_id;
13 -- Region 5
14 SELECT motif_id, count(*) FROM motif_locations WHERE start > 200 - (char_length(motif_id) / 2) AND start <= 250 - (char_length(motif_id) / 2)
15 GROUP BY motif_id;
16 -- Region 6
17 SELECT motif_id, count(*) FROM motif_locations WHERE start > 250 - (char_length(motif_id) / 2) AND start <= 300 - (char_length(motif_id) / 2)
18 GROUP BY motif_id;
19 -- Region 7
20 SELECT motif_id, count(*) FROM motif_locations WHERE start > 300 - (char_length(motif_id) / 2) AND start <= 350 - (char_length(motif_id) / 2)
21 GROUP BY motif_id;
22 -- Region 6_2
23 SELECT motif_id, count(*) FROM motif_locations WHERE start > 350 - (char_length(motif_id) / 2) AND start <= 400
24 GROUP BY motif_id;
25 -- Region 6_1
26 SELECT motif_id, count(*) FROM motif_locations WHERE start > 250 - (char_length(motif_id) / 2) AND start <= 262 - (char_length(motif_id) / 2)
27 GROUP BY motif_id;
28 -- Region 6_2
29 SELECT motif_id, count(*) FROM motif_locations WHERE start > 262 - (char_length(motif_id) / 2) AND start <= 282 - (char_length(motif_id) / 2)
30 GROUP BY motif_id;
31 -- Region 6_3
32 SELECT motif_id, count(*) FROM motif_locations WHERE start > 282 - (char_length(motif_id) / 2) AND start <= 300 - (char_length(motif_id) / 2)
33 GROUP BY motif_id;

```

A snapshot from “motif\_locations” table which stores FIMO output and SQL queries used for getting motif occurrences on the 8 regions around the splice acceptor sites

a.								
	1	MA0528.1	356	52361541	376	52361561	+	0.00000000000595000
	2	MA0528.1	317	10642286	337	10642306	-	0.000000000002870000
	3	MA0528.1	9	201874860	29	201874880	+	0.000000000002870000
	4	MA0528.1	329	201874860	349	201874880	+	0.000000000002870000
	5	MA0528.1	335	1641766	355	1641786	+	0.000000000002870000
	6	MA0528.1	335	1708998	355	1708918	+	0.000000000002870000
	7	MA0528.1	359	52361538	379	52361558	+	0.000000000004020000
	8	MA0528.1	332	1708901	352	1708921	+	0.000000000007890000
	9	MA0528.1	332	1641769	352	1641789	+	0.000000000007890000
	10	MA0528.1	233	20342364	253	20342384	-	0.000000000008290000

b.								
	1	670491	chr1	91661	91662			
	2	670492	chr1	169295	169296			
	3	670493	chr1	498477	498478			
	4	670494	chr1	769743	769744			
	5	670495	chr1	773130	773131			
	6	670496	chr1	781909	781910			
	7	670497	chr1	849444	849445			
	8	670498	chr1	852647	852648			
	9	670499	chr1	914759	914760			
	10	670500	chr1	918144	918145			

c.	<i>-- Retrieve motif occurrences on branch points</i>								
	1								
	2	<b>SELECT</b>							
	3	<b>bpounds.id,</b>							
	4	<b>motif_id,</b>							
	5	<b>count(*)</b>							
	6	<b>FROM</b>							
	7	<b>motif_locations ml,</b>							
	8	<b>genome_wide_branch_point_locations bpoints</b>							
	9	<b>WHERE ml.p_value &lt;= 0.0001 AND ml.real_start &lt;= bpoints.chr_start AND ml.real_stop &gt;= bpoints.chr_end</b>							
	10	<b>GROUP BY bpoints.id, motif_id;</b>							

a. A snapshot from “motif\_locations” table which stores FIMO output, b. A snapshot from “genome\_wide\_branch\_point\_locations” table which stores branch point locations published in “Genome-wide discovery of human splicing branchpoints” and SQL queries used for getting motif occurrences on the branchpoints.