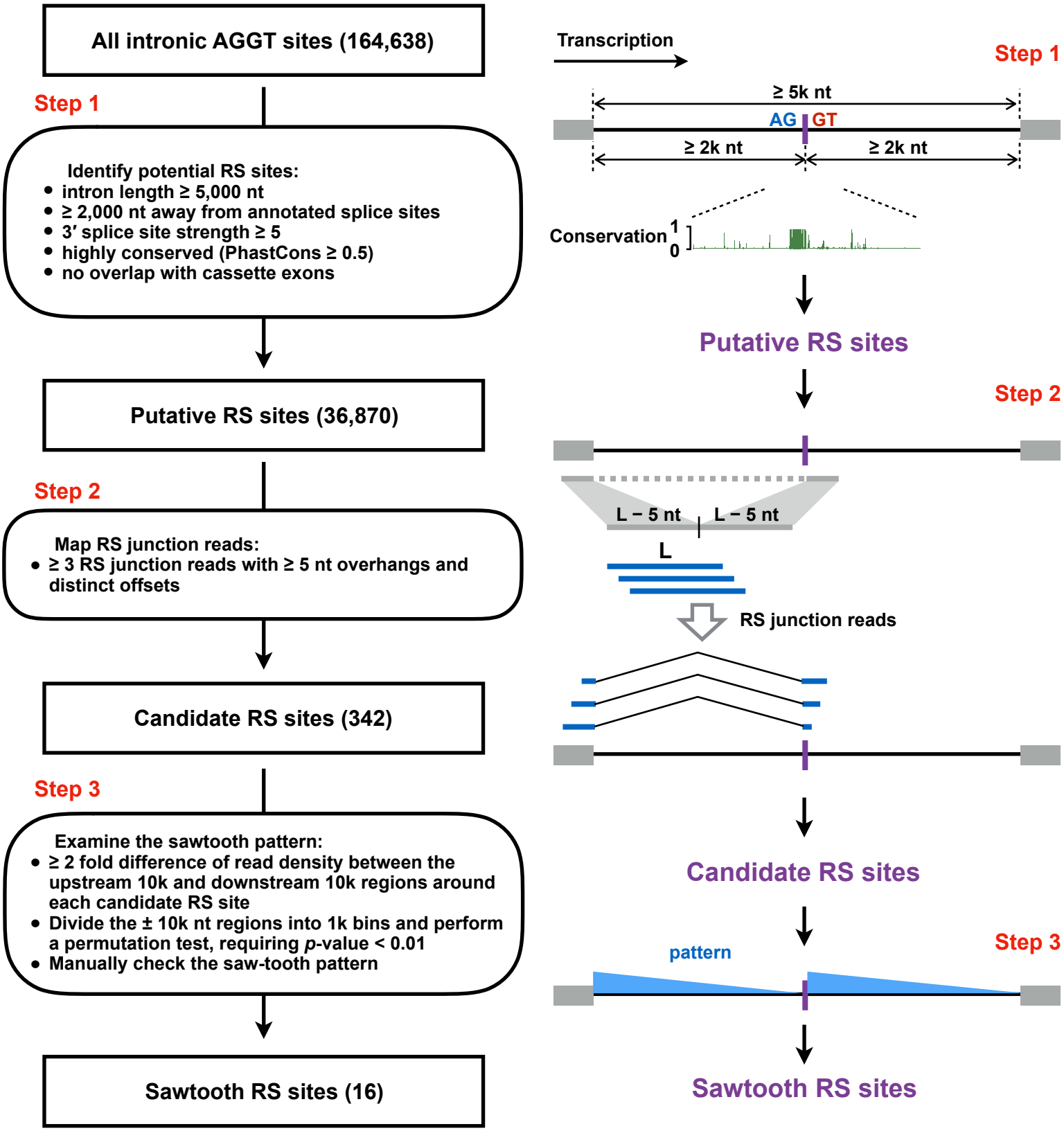
All intronic AGGT sites (164,638)
Step 1
Identify potential RS sites:
• intron length ≥ 5,000 nt
• ≥ 2,000 nt away from annotated splice sites
• 3′ splice site strength ≥ 5
• highly conserved (PhastCons ≥ 0.5)
• no overlap with cassette exons
Putative RS sites (36,870)
Step 2
Map RS junction reads:
• ≥ 3 RS junction reads with ≥ 5 nt overhangs and distinct offsets
Candidate RS sites (342)
Step 3
Examine the sawtooth pattern:
• ≥ 2 fold difference of read density between the upstream 10k and downstream 10k regions around each candidate RS site
• Divide the ± 10k nt regions into 1k bins and perform a permutation test, requiring p-value < 0.01
• Manually check the saw-tooth pattern
Sawtooth RS sites (16)
Transcription
Step 1
≥ 5k nt
AG
GT
≥ 2k nt
≥ 2k nt
Conservation
1
0
Putative RS sites
Step 2
L − 5 nt
L − 5 nt
L
RS junction reads
Candidate RS sites
Step 3
pattern
Sawtooth RS sites