CLAIMS

What is claimed is:

[Claim 1]. A system comprising:

at least one processor; and

a non-transitory computer readable storage medium comprising instructions that, when executed by the at least one processor, cause the system to:

identify a target nucleotide sequence comprising a target nucleotide and contextual nucleotides;

provide the target nucleotide sequence as an input into an atrous convolutional neural network trained to generate splice site predictions for input nucleotide sequences based on feature extractions, wherein the atrous convolutional neural network comprises groups of residual blocks parameterized by a number of convolutional filters in residual blocks of each respective group of residual blocks; and

generate, utilizing the atrous convolutional neural network to perform atrous convolutions to extract learned features representing a splicing pattern of the target nucleotide sequence, a splice site prediction for the target nucleotide based on the contextual nucleotides within the target nucleotide sequence.

[Claim 2]. The system of claim 1, wherein the target nucleotide comprises a variant nucleotide.

[Claim 3]. The system of claim 1, wherein the contextual nucleotides comprise at least 200 upstream nucleotides flanking the target nucleotide and at least 200 downstream nucleotides flanking the target nucleotide.

[Claim 4]. The system of claim 1, wherein the contextual nucleotides comprise at least 5000 upstream nucleotides flanking the target nucleotide and at least 5000 downstream nucleotides flanking the target nucleotide.

[Claim 5]. The system of claim 1, further comprising instructions that, when executed by the at least one processor, cause the system to generate, by the atrous convolutional neural network, the splice site prediction for the target nucleotide by generating a splice site score for the target nucleotide and a non-splicing site score for the target nucleotide based on the contextual nucleotides within the target nucleotide sequence.

[Claim 6]. The system of claim 5, further comprising instructions that, when executed by the at least one processor, cause the system to identify a highest score among the splice site score for the target nucleotide and the non-splicing site score for the target nucleotide.

[Claim 7]. The system of claim 1, wherein the groups of residual blocks are further parameterized by a convolution window size of the residual blocks thereof and an atrous convolution rate of the residual blocks thereof.

[Claim 8]. The system of claim 7, wherein the atrous convolution rate comprises atrous convolutions that conserve partial convolution calculations for reuse as flanking nucleotides of the contextual nucleotides are processed by the atrous convolutional neural network.

[Claim 9]. The system of claim 1, wherein the groups of residual blocks are further parameterized by a number of residual blocks, a number of skip connections, and a number of residual connections.

[Claim 10]. The system of claim 1, wherein each group of residual blocks produces an intermediate output by processing a preceding input, wherein dimensionality of the intermediate output is $(I-[\{(W-1)*D\}*A])\times N$, where:

I is dimensionality of the preceding input;

W is convolution window size of the residual blocks of each respective group of residual blocks;

D is atrous convolution rate of the residual blocks of each respective group of residual blocks;

A is a number of atrous convolution layers in each respective group of residual blocks; and

N is a number of convolution filters in the residual blocks of each respective group of residual blocks.

[Claim 11]. A non-transitory computer readable storage medium storing instructions that, when executed by at least one processor, cause a system to:

identify a target nucleotide sequence comprising a target nucleotide and contextual nucleotides;

provide the target nucleotide sequence as an input into an atrous convolutional neural network trained to generate splice site predictions for input nucleotide sequences based on feature extractions, wherein the atrous convolutional neural network comprises groups of residual blocks parameterized by a number of convolutional filters in residual blocks of each respective group of residual blocks; and

generate, utilizing the atrous convolutional neural network to perform atrous convolutions to extract learned features representing a splicing pattern of the target nucleotide sequence, a splice site prediction for the target nucleotide based on the contextual nucleotides within the target nucleotide sequence.

[Claim 12]. The non-transitory computer readable storage medium of claim 11, wherein the target nucleotide comprises a variant nucleotide.

[Claim 13]. The non-transitory computer readable storage medium of claim 11, wherein the contextual nucleotides comprise at least 200 upstream nucleotides flanking the target nucleotide and at least 200 downstream nucleotides flanking the target nucleotide.

[Claim 14]. The non-transitory computer readable storage medium of claim 11, further storing instructions that, when executed by the at least one processor, cause the system to generate, by the atrous convolutional neural network, the splice site prediction for the target nucleotide by generating a splice site score for the target nucleotide and a non-splicing site score for the target nucleotide based on the contextual nucleotides within the target nucleotide sequence.

[Claim 15]. The non-transitory computer readable storage medium of claim 14, further storing instructions that, when executed by the at least one processor, cause the system to identify a highest score among the splice site score for the target nucleotide and the non-splicing site score for the target nucleotide.

[Claim 16]. A computer-implemented method comprising:

identifying a target nucleotide sequence comprising a target nucleotide and contextual nucleotides;

providing the target nucleotide sequence as an input into an atrous convolutional neural network trained to generate splice site predictions for input nucleotide sequences based on feature extractions, wherein the atrous convolutional neural network comprises groups of residual blocks parameterized by a number of convolutional filters in residual blocks of each respective group of residual blocks; and

generating, utilizing the atrous convolutional neural network to perform atrous convolutions to extract learned features representing a splicing pattern of the target nucleotide sequence, a splice site prediction for the target

nucleotide based on the contextual nucleotides within the target nucleotide sequence.

[Claim 17]. The computer-implemented method of claim 16, wherein the groups of residual blocks are further parameterized by a convolution window size of the residual blocks thereof and an atrous convolution rate of the residual blocks thereof.

[Claim 18]. The computer-implemented method of claim 17, wherein the atrous convolution rate comprises atrous convolutions that conserve partial convolution calculations for reuse as flanking nucleotides of the contextual nucleotides are processed by the atrous convolutional neural network.

[Claim 19]. The computer-implemented method of claim 16, wherein the groups of residual blocks are further parameterized by a number of residual blocks, a number of skip connections, and a number of residual connections.

[Claim 20]. The computer-implemented method of claim 16, wherein each group of residual blocks produces an intermediate output by processing a preceding input, wherein dimensionality of the intermediate output is (I-[{(W-1) * D} * A]) x N, where:

I is dimensionality of the preceding input;

W is convolution window size of the residual blocks of each respective group of residual blocks;

D is atrous convolution rate of the residual blocks of each respective group of residual blocks;

A is a number of atrous convolution layers in each respective group of residual blocks; and

N is a number of convolution filters in the residual blocks of each respective group of residual blocks.