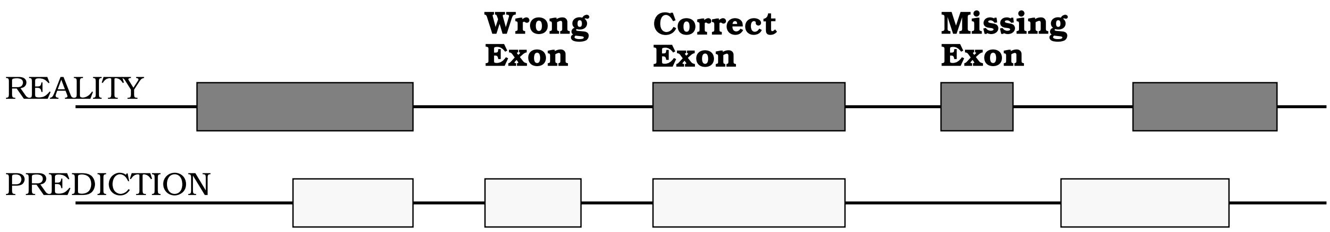
2 Measures of Prediction Accuracy

2.2 Exon Level



An **exon** is assumed to be **correctly predicted** if the overlap between actual and predicted exon is greater or equal than a given threshold α**.**

$$Sn = \frac{\text{number of Correct Exons}}{\text{number of Actual Exons}}$$ **Sensitivity**

$$Sn = \frac{\text{number of Correct Exons}}{\text{number of Predicted Exons}}$$ **Specificity**

$$ME = \frac{\text{number of Missing Exons}}{\text{number of Actual Exons}}$$ **(Sensitivity)**

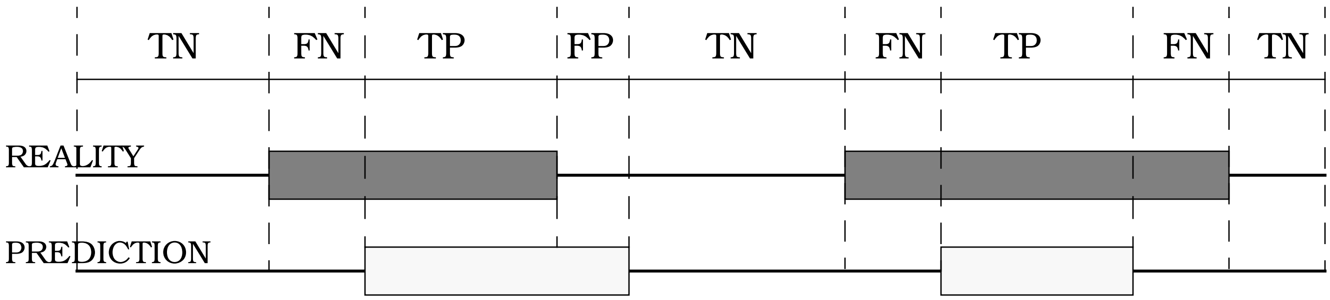
$$WE = \frac{\text{number of Wrong Exons}}{\text{number of Predicted Exons}}$$ **(Specificity)**

2.3 Protein Level

Find the **global alignment** between the protein sequence encoded by the **actual** gene and the protein sequence encoded by the **predicted** gene and compute the **% similarity**

2 Measures of Prediction Accuracy

2.1 Nucleotide Level



REALITY

		coding	no coding	
PREDICTION	coding	TP	FP	TP+FP
	no coding	FN	TN	FN+TN
		TP+FN	TF+TN	

$$Sn = \frac{TP}{TP + FN} = P\langle C_p | C_r \rangle$$

Sensitivity

$$Sp = \frac{TP}{TP + FP} = P\langle C_r | C_p \rangle$$

Specificity

$$CC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}}$$

Correlation Coefficient

CC depends on $P\langle C_p | C_r \rangle, P\langle C_r | C_p \rangle, P\langle \neg C_p | \neg C_r \rangle, P\langle \neg C_r | \neg C_p \rangle$

$$ACP = \frac{1}{4}\left[\frac{TP}{TP + FN} + \frac{TP}{TP + FP} + \frac{TN}{TN + FP} + \frac{TN}{TN + FN}\right]$$

$$AC = (ACP - 0.5) \times 2$$

Approximate Correlation