Colour Blindness- Assignment:4

E0 259 Data Analytics

Hemanth Kongara MTech AI 18086

Summary:

- Stored all the inputs using appropriate data structures to minimize the memory required
- Calculated the ranks for all the rows and stored them in a numpy array (i.e. Δ=1)
- All 'N''s in reads are replaced with 'A' before any processing
- 'get_complimentary' function returns the reverse compliment of the input read
- 'update_band' function takes current band and next char (in reverse sequence of input read) as input and outputs the updated band
- 'get_starting_band' function is used to get the initial band to start with, which happens at starting of the read and after a mismatch
- 'get_matching_info' function takes read as input and outputs the final band to search & align and offset which depends on the index where the final band is found. The exit conditions are length of band less than 3 or mismatches greater than 2.
- 'get_starting_index' function takes the final band, offset and read as inputs and outputs all the possible starting indexed with mismatches less than are equal to two.
- 'find_gene' function takes read as input and outputs the 2*6 numpy array with information of corresponding matched exon
- All the reads are passed to 'find_gene' function and all the exon information obtained is added to a variable exon_count which is initialized with all zeros
- 'generate_prob' function takes model information and exon_count as input and outputs the probability of generating these counts for the input model information

Results:

	Exon 1	**Exon 2**	**Exon 3**	**Exon 4**	**Exon 5**	**Exon 6**
Red Exons	**179.5**	**152.5**	**102**	**175.5**	**319.5**	**404.5**
Green Exons	**195.5**	**224.5**	**148**	**152.5**	**381.5**	**403.5**

The most likely configuration that lead to colour-blindness is "Fraction on Red Exons 2,3,4,5 is 33%,33%,100%,100%".