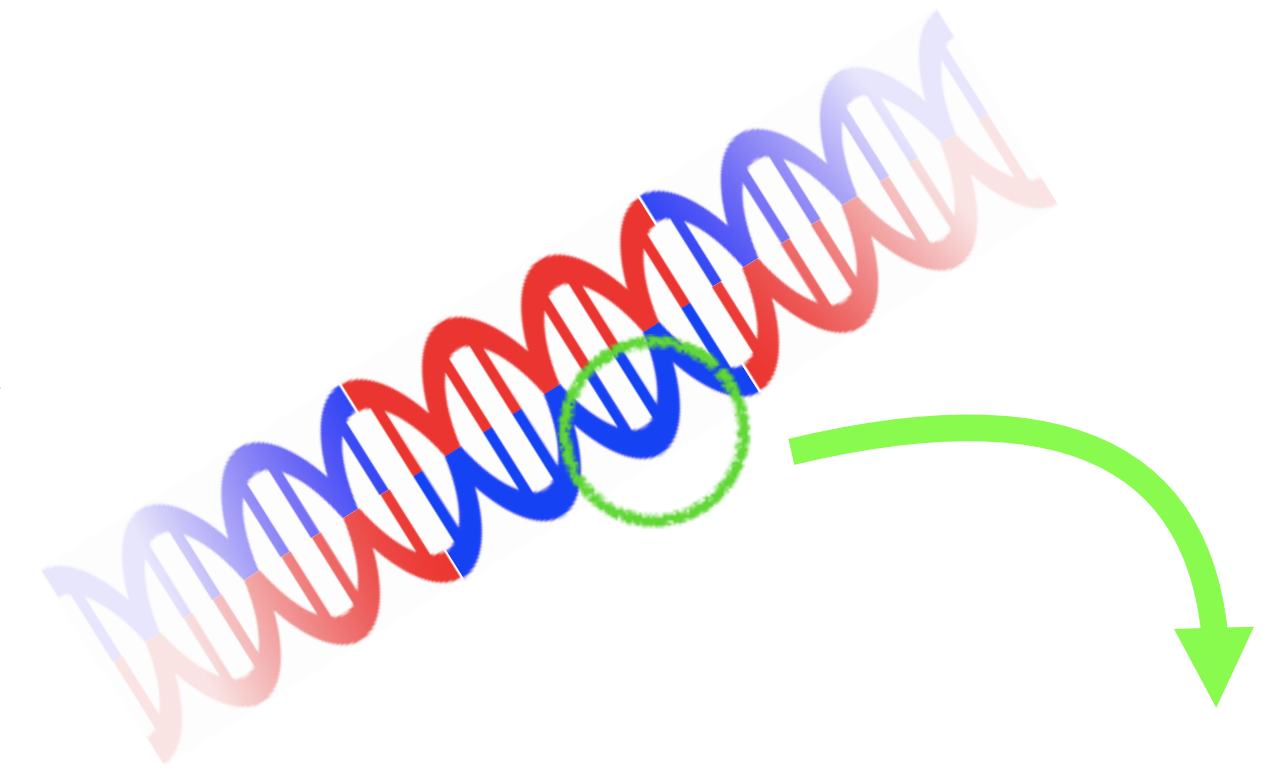
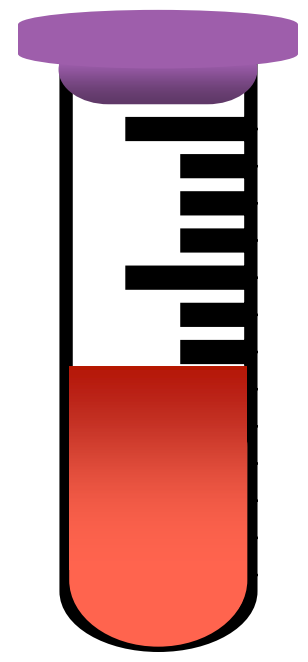


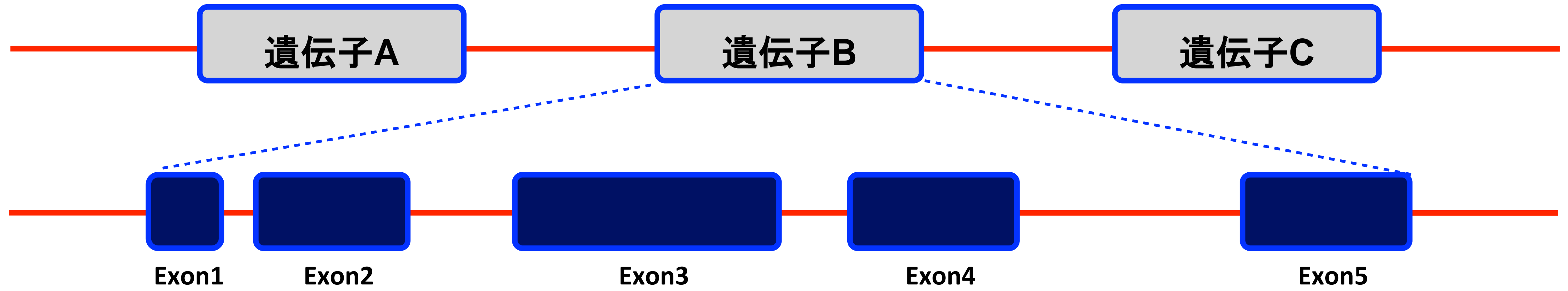
# ～NGS法による解析の流れ～

イルミナシーケンサー、キャプチャー法



### 1. DNA の抽出

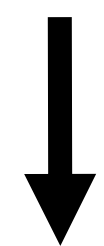
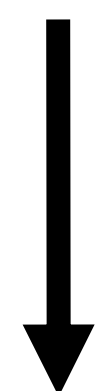
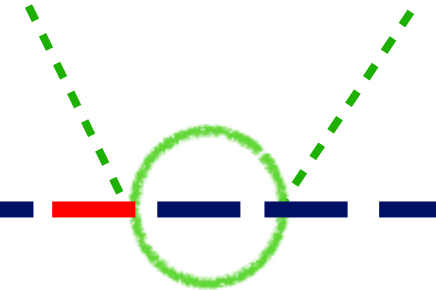
血液検体からゲノムDNA を抽出します。  
白血球の核より抽出しますので、血液量および白血球数が少ない場合は、検査実施には不十分な可能性もあります。



### 2. DNA の断片化

DNA はエクソン（青い四角）とイントロン（赤線）が組み合わさっており、エクソン部分がヒトの体をつくるタンパクをコードしています。  
抽出したDNA（全長約3Gbp）を200bp 前後に細切れにします。

約200塩基



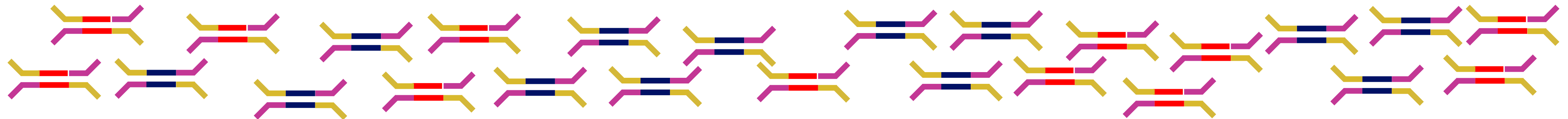
# 断片化されたDNA

アダプター



## 3. ライブラリの調整

断片化したDNA にアダプターと呼ばれる配列をくっつけて NGS で解析できる形にします。



## 4. DNA のキャプチャー

このままでは解析に不要な遺伝子やイントロン部分も含まれるため、配列に特異的なプローブを用いて疾患に関連する領域を物理的に集める作業をします。



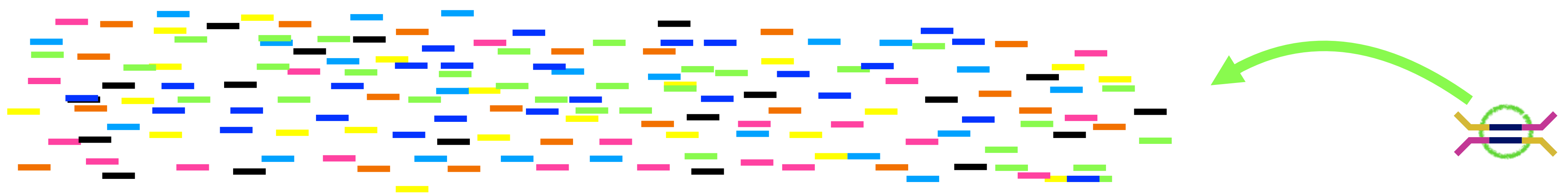
## 5. NGS での解析

ライブラリをロードして解析を開始します。

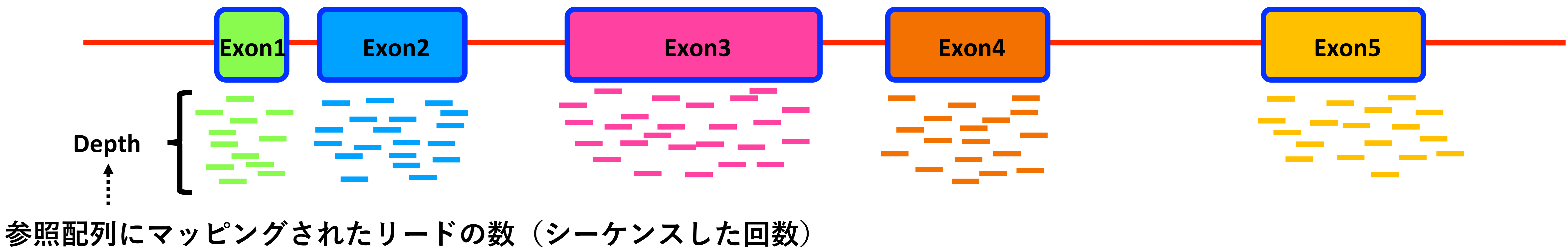
\* NGS の原理については省略



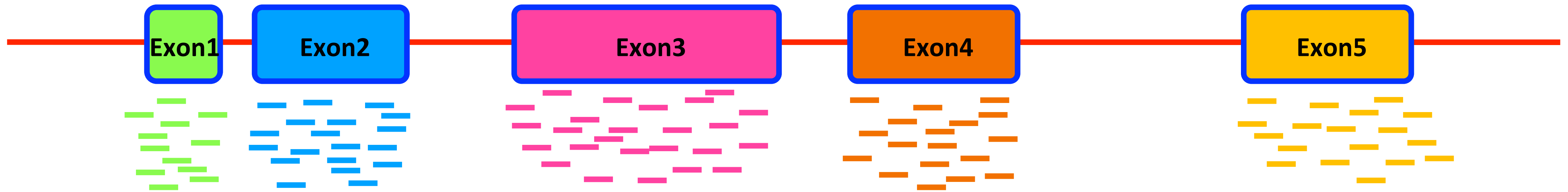
➤ NGS で解析後、断片化されたDNA の塩基配列情報（緑の丸枠部分）が出力されます。この1本1本をリードと言います。



6. データ解析（マッピング/アライメント）  
得られたリードの塩基配列情報をもとに、基準となるhg38 の参照配列にマッピングして位置を同定します。



参照配列にマッピングされたリードの数（シーケンスした回数）



7. データ解析 (バリエントコール)  
位置が決まったリードから配列の変異を抽出します。

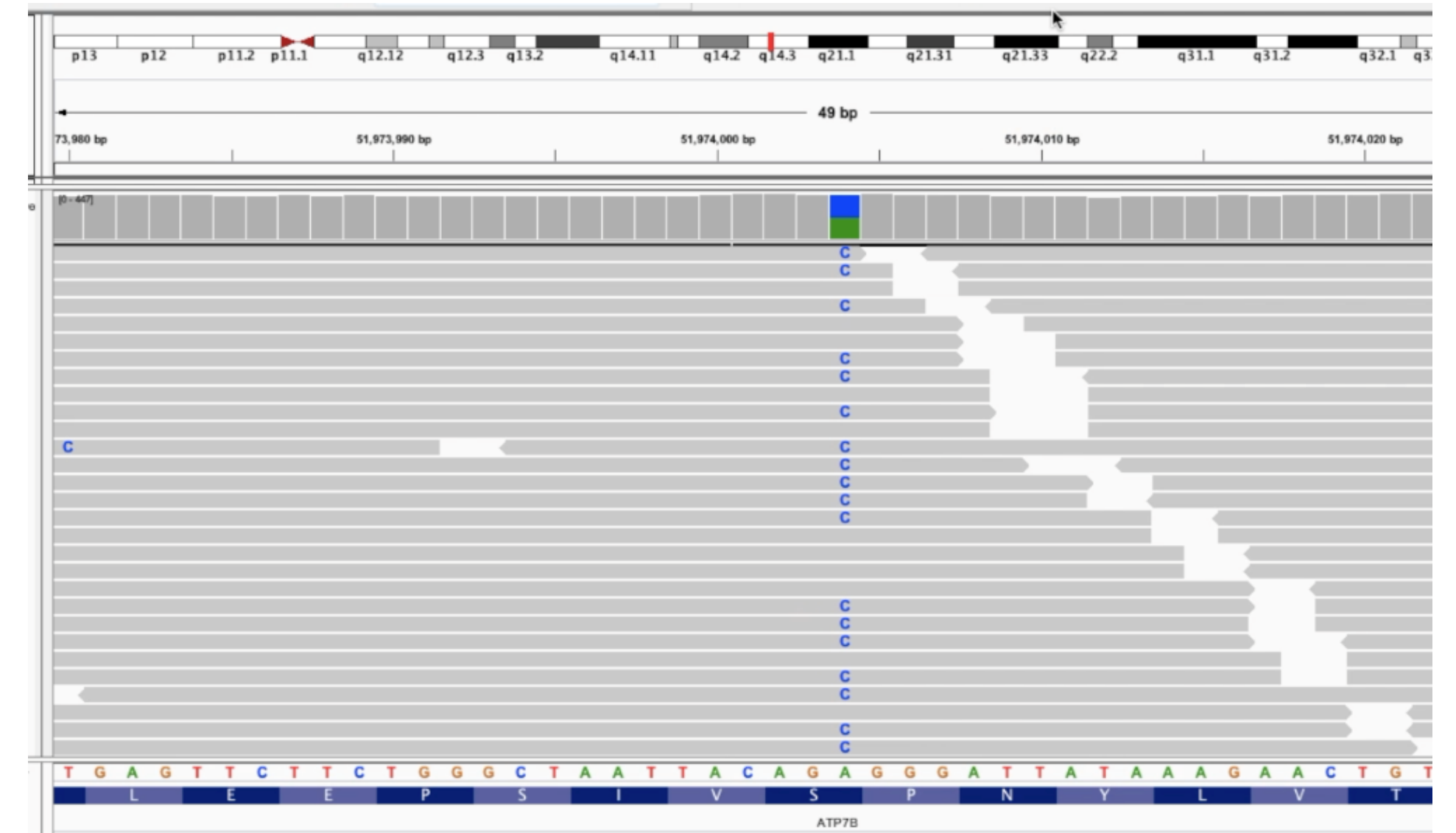
8. データ解析 (可視化)  
マッピングされた塩基配列情報をIGV やUCSC GenomeBrowser を使用することで、データの確認・解釈をします。

参照配列  
(hg38)

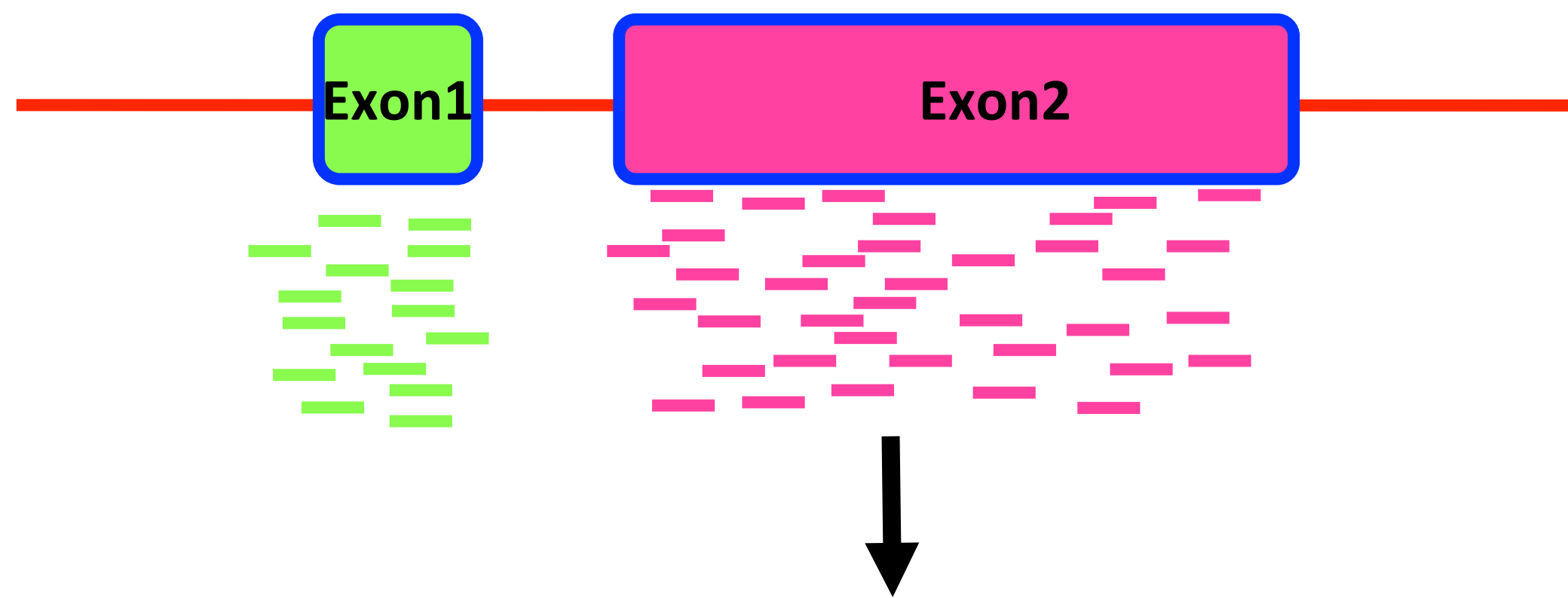
ATGCCGTACGTTACGTA**CT**CTTTTCGTATAGCT

リード 1	ACGTTACGTA <b>CT</b> CTTTTCGTATAG
リード 2	GTACGTTACGTA <b>C</b> CTTTTCGTAT
リード 3	TTACGTA <b>CT</b> CTTTTCGTATAGCT
リード 4	CGTTACGTA <b>C</b> CTTTTCGTATAGC
リード 5	CCGTACGTTACGTA <b>C</b> CTCTT
リード 6	ATGCCGTACGTTACGTA <b>CT</b>
リード 7	ACGTA <b>CT</b> CTTTTCGTATAGCT . . .
⋮	⋮

➤ 黄色の枠部分が参照配列Gに対して、データではGとCが半分になっているため、ヘテロの変異が疑われます。右はIGVでのデータ表示の例です。



➤ データの解釈が難しいパターン1 (繰り返し配列)



参照配列  
(hg38)

GT**TA**AGTCGTCGTCAGTCGTCGTCAGTCGTCGTCAGTCGTCGTCAGTCGTCGTCAGTCGTCAGTCGTCAGTCGTCAGTCGTCAGTCGTCGTCAGTCGTCAGTCGTC**AA**CCA

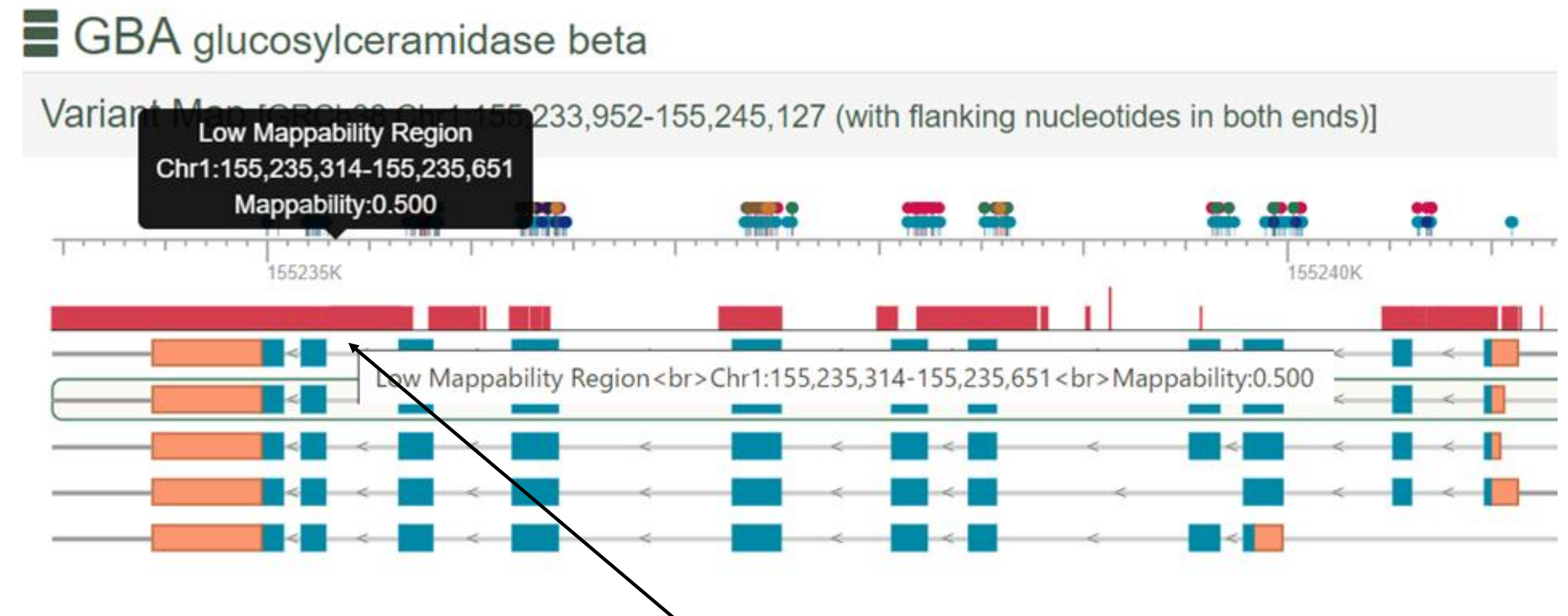
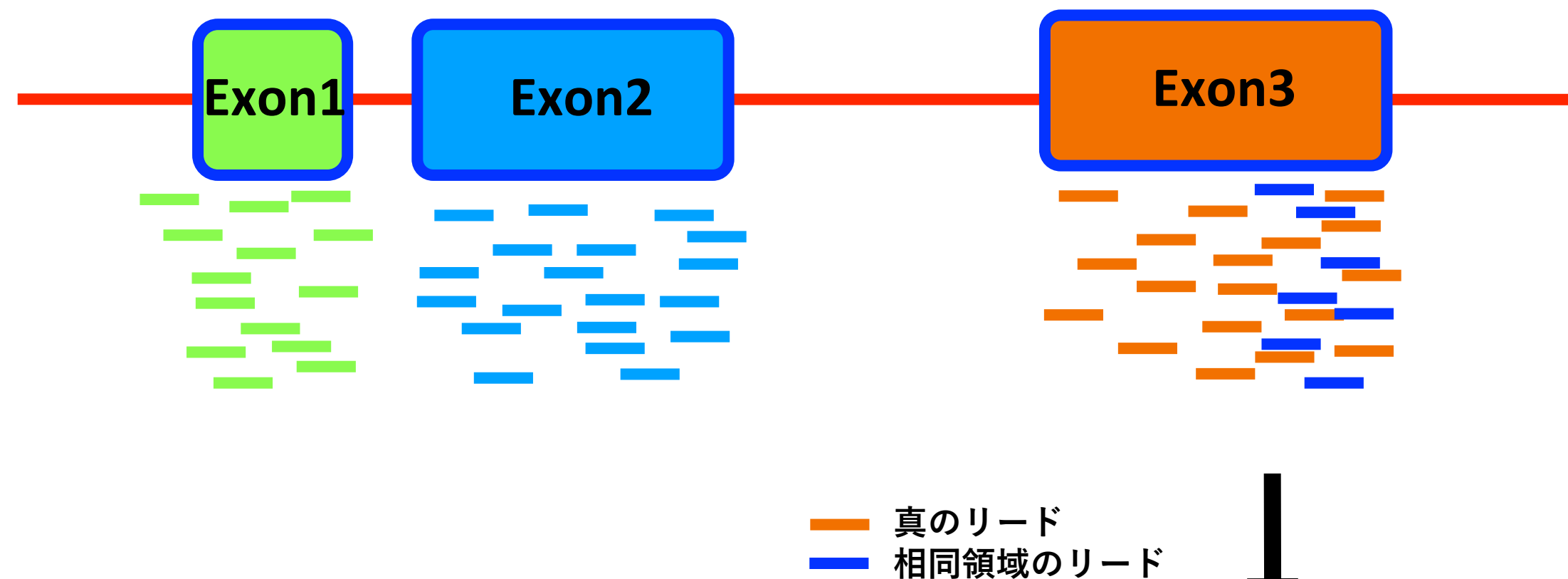
リード 1	GT <b>TA</b> AGTCGTCGTCAGTCGTCGTCAGTCGTCGTC	
リード 2		GTCGTCAGTCGTCGTCAGTCGTCGTCGTC <b>AA</b> CCA
リード 3		GTCAGTCGTCGTCAGTCGTCGTCAGTCGTCGTCAGTCGTC
リード 4	<u>AGTCGTCGTCAGTCGTCGTCAGTCGTCGTC</u>	
リード 5		AGTCGTCGTCAGTCGTCGTCAGTCGTCAGTCGTCGTCAGTCGTCGTC
リード 6	<u>AGTCGTCGTCAGTCGTCGTCAGTCGTCGTC</u>	
リード 7		CGTCGTCAGTCGTCAGTCGTCGTCAGTCGTCAGTCGTCGTC
		⋮

リード4、6は同じ配列なのに違う位置にあるけど、どっちが正解??

➤ リード1と2は繰り返し配列以外を認識して、正しくマッピングされますが、それ以外の配列は、AGTCGTCGTCが繰り返す配列(黄色枠)のため、正しくマッピングされないことがあります。



## データの解釈が難しいパターン2 (相同領域がある場合)



例としてGBAのように、相同領域やPseudogeneがある遺伝子 (Mappabilityが1未満) だと、正しい位置にマッピングされないことがあります。

参照配列  
(hg38)

GTATGCTCGGTAGGATAAAGTCGTTTGC**G**ATGCATGCATCGACGTCAAGTAGA

リード 1	GTATGCTCGGTAGGATAAAGTCGTTTGC <b>G</b> ATG
リード 2	TTTGC <b>G</b> ATGCATGCATCGACGTCAAGTAGA
リード 3	TCGGTAGGATAAAGTCGTTTGC <b>C</b> TGCATGCATCGACGTCA
リード 4	GGATAAAGTCGTTTGC <b>C</b> TGCATGCATCGACGTCAAGTAGA
リード 5	TCGTTTGC <b>G</b> ATGCATGCATCGACGTCAAGTAGA
リード 6	GTATGCTCGGTAGGATAAAGTCGTTTGC <b>G</b> ATGCATGCA
⋮	⋮

➤ エクソン3において、相同領域がある場合、正しい位置にマッピングされないため、別の領域のリードが混じる可能性があります。リード3と4で、参照配列Aに対してCが検出されていますが、これが真のリードか相同領域のリードか区別できない可能性があります。