

Shazya Audrea Taufik - 13522063 | Zahira Dina Amalia - 13522085

## **Topik:** 1. Analisis Sekuens DNA/RNA

### **Judul:** Analisis Komparatif Varian Sekuens Gen BRCA1 Menggunakan Algoritma Pencocokan Sekuens

#### **1. Latar Belakang**

##### ***Biological Background:***

BRCA1 (Breast Cancer 1) adalah gen penting dalam perbaikan DNA dan pengendalian siklus sel. Mutasi pada gen ini berhubungan erat dengan risiko kanker payudara dan ovarium hereditas. Deteksi dini varian patogenik BRCA1 sangat krusial dalam diagnosis dan penanganan risiko kanker.

##### ***Computational Biology Background:***

Dalam penelitian bioinformatika, analisis sekuens dan klasifikasi varian genetik telah menjadi metode yang umum digunakan untuk mengidentifikasi mutasi. Algoritma seperti Smith-Waterman dan Needleman-Wunsch memungkinkan perbandingan akurat antara sekuens referensi dan sampel. Namun, integrasi pipeline otomatis yang menghubungkan hasil pencocokan dengan klasifikasi klinis (misalnya dari ClinVar) masih jarang diterapkan.

##### ***Problem Formulation & Objectives:***

- Membangun pipeline otomatis untuk mengidentifikasi dan mengklasifikasi varian pada sekuens BRCA1.
- Menerapkan algoritma pencocokan lokal dan global untuk mengidentifikasi mutasi spesifik dan perbandingan sekuens secara keseluruhan.
- Mengintegrasikan hasil pencocokan dengan database ClinVar untuk klasifikasi klinis varian.

#### **2. Metode / Algoritma**

##### ***Metode Komputasi:***

- **Pencocokan Lokal:** Algoritma Smith-Waterman
- **Pencocokan Global:** Algoritma Needleman-Wunsch
- **Multiple Sequence Alignment:** Menggunakan MUSCLE untuk analisis antar populasi/spesies
- **Klasifikasi Varian:** Mapping ke database ClinVar untuk signifikansi klinis

##### ***Alasan Pemilihan:***

- Smith-Waterman unggul dalam mengidentifikasi mutasi lokal (substitusi, insersi, delesi)
- Needleman-Wunsch sesuai untuk membandingkan keseluruhan sekuens

- MUSCLE efisien dalam MSA untuk studi evolusi/populasi

***Flowchart Singkat:***

- 1) Ambil sekuens referensi BRCA1 dari NCBI GenBank
- 2) Ambil sekuens sampel dari pengguna/input
- 3) Lakukan pencocokan lokal (Smith-Waterman)
- 4) Lakukan pencocokan global (Needleman-Wunsch)
- 5) Identifikasi titik mutasi
- 6) Klasifikasi varian berdasarkan ClinVar
- 7) Tampilkan hasil dengan skor alignment dan klasifikasi klinis

### 3. Dataset

- **Referensi:** NCBI GenBank (BRCA1), ClinVar (varian BRCA1)
- **Format:** FASTA (sekuens DNA), JSON/TSV (ClinVar annotations)
- **Ukuran Data:** ~100-200kb per sekuens
- **Karakteristik:** Nukleotida A, T, C, G dengan lokasi varian teranotasi di ClinVar

***Contoh Input:***

```
>Sample_Pasien_123
ATGGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAAATGTCATTAATGCTATGCAGAAAATCTTGAGTGTC
CCATCTGTCTGGAGTTGATCAAGG...
```

***Contoh Output:***

```
Posisi 185-185: Referensi (A) -> Sampel (C)
Klasifikasi Varian: c.185A>C (p.Gln62Pro)
Signifikansi Klinis: Patogenik (ClinVar ID: SCV000012345)
Alignment Score: 92.7
```

### 4. Interface Design

***Desain Awal UI (mockup):***

- Input: Form unggah sekuens DNA (FASTA)
- Tombol: "Analisis"
- Output:
  - Tabel perbandingan sekuens
  - Mutasi terdeteksi dan klasifikasi klinis
  - Skor alignment

***Interaksi Pengguna:***

- Unggah file sekuens
- Klik tombol "Analyze Sequence"

- Lihat hasil dalam tabel “Analysis Result”

The screenshot displays the BRCA1 Sequence Analyzer web application. The interface is divided into two main panels: 'Analysis Parameters' on the left and 'Analysis Results' on the right.

**Analysis Parameters Panel:**

- Header:** BRCA1 Sequence Analyzer (with a green logo).
- Section: Analysis Parameters** (Configure the parameters for your BRCA1 sequence analysis).
- Email (for NCBI Entrez queries):** Input field with 'your.email@example.com'.
- Reference Sequence Accession:** Input field with 'NM\_007294.4'.
- Alignment Type:** Dropdown menu set to 'Global'.
- Variant Annotation Source:** Dropdown menu set to 'ClinVar'.
- Population (optional):** Input field with 'e.g., European, African, Asian'.
- Buttons:** 'Upload FASTA File' and 'Paste Sequence'.
- File Upload Area:** A dashed box with a green upload icon and the text 'Drag and drop your FASTA file here, or' and a 'Browse Files' button.
- Analyze Sequence:** A large green button at the bottom.

**Analysis Results Panel:**

- Section: Analysis Results** (Detailed analysis of sample Sample\_123).
- Export/Share:** Buttons for 'JSON', 'CSV', 'Share', and 'Download Report'.
- Metadata Table:**

Sample ID	Reference ID	Sample Length
Sample_123	NM_007294.4	854 bp

Alignment Score	Processing Time	Analysis Date
726.50	4.82 seconds	April 22, 2025 10:10:47 AM
- Variant Summary:**
  - Variant Types:**

Variant Type	Count
SNP	4
Indel	1
  - Clinical Significance:**

Clinical Significance	Count
Pathogenic	1
Benign	1
Likely pathogenic	1
Uncertain significance	1
Likely benign	1
- Key Findings:**
  - Pathogenic Variants Detected:** This sample contains 2 pathogenic or likely pathogenic variant(s) that may be clinically significant.
    - c.185A>C - Pathogenic
    - c.356del - Likely pathogenic

## 5. Evaluation & Testing

### Evaluasi Sistem:

- **Validasi Internal:** Bandingkan hasil identifikasi mutasi dengan sekuens ground truth dari ClinVar
- **Evaluasi Kinerja:**
  - Akurasi (>98%)
  - Waktu proses (rata-rata < 5 detik per sekuens)
  - Kemampuan batch processing (uji 100 sekuens sekaligus)

### Strategi Verifikasi Biologis:

- Cross-reference hasil klasifikasi dengan studi literatur atau anotasi ClinVar terkini
- Bandingkan hasil klasifikasi dengan hasil laboratorium atau studi pasien jika tersedia

### Test Plan:

- Uji pada 5 sekuens sintetis (dengan varian diketahui)
- Uji pada 50 sekuens nyata dari studi publik
- Analisis kesalahan pencocokan atau klasifikasi varian

## 6. Reference dan Perbandingan

- NCBI GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>
- ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>

Perbandingan dari jurnal-jurnal yang sudah ada dapat terlihat dari tabel berikut.

Jurnal	Metode	Pembeda Jurnal	Pembeda Proposal
<a href="#">Jarhelle et al., 2019 (Scientific Reports - Nature)</a>	NGS & 94 gen analisis	Tidak mengintegrasikan klasifikasi otomatis dari ClinVar	Output otomatis beserta klasifikasi klinis lengkap
<a href="#">Damiola et al., 2015: "Mutation analysis of PALB2 gene in French breast cancer families"</a>	Identifikasi mutasi PALB2	Fokus pada PALB2, bukan BRCA1	Fokus spesifik pada BRCA1
<a href="#">Lindor et al., 2012 (Breast Cancer Research): "A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS)"</a>	Analisis konservasi struktural	Tidak otomatis, tidak mendukung batch	Pipeline otomatis, bisa proses batch
<a href="#">Goldgar et al., 2004 (American Journal of Human Genetics): "Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance: Application to BRCA1 and BRCA2"</a>	Multifactorial likelihood	Tidak menggunakan pencocokan sekuens	Menggunakan Smith-Waterman dan Needleman-Wunsch