# MTB Genomic Data Request Checklist

**Purpose:** This checklist standardizes genomic data requests across collaborating sites (Ulm, Heidelberg, Freiburg, Tuebingen) to ensure comprehensive molecular profiling for personalized oncology treatment decisions.

## 1. Patient & Sample Information

- ☐ Patient ID (anonymized/pseudonymized)
- ☐ Sample ID(s)
- ☐ Sample type (tumor tissue, blood, circulating tumor DNA, etc.)
- ☐ Tumor type/cancer diagnosis
- ☐ Sample collection date
- ☐ Clinical stage at time of sample collection

## 2. Sequencing & Technical Information

### 2.1 Sequencing Platform Details

- ☐ Sequencing platform (e.g., Illumina NovaSeq, HiSeq 4000, NextSeq)
- ☐ Library preparation kit (manufacturer and product name)
- ☐ Kit type (e.g., Amplicon, Targeted Resequencing, Whole Exome, RNA-Seq)
- ☐ Gene panel name/version (if targeted sequencing)
- ☐ Target coverage depth (mean/median coverage)

### 2.2 Bioinformatics Pipeline

- ☐ Pipeline software and version (e.g., CLC Genomics Workbench 23.0.3)
- ☐ Reference genome build (HG19/GRCh37 or HG38/GRCh38)
- ☐ Variant calling algorithm/parameters
- ☐ Filtering criteria (e.g., VAF threshold, coverage requirements)

## 3. Genomic Alterations Data

### 3.1 Single Nucleotide Variants (SNVs) & Indels

- ☐ Complete variant list (VCF file or tabular format)

Required fields for each variant:

- Gene name
- Chromosome position (chr:pos)
- Reference and alternate alleles
- Variant consequence (e.g., missense, nonsense, frameshift)
- Amino acid change (if applicable)
- Variant allele frequency (VAF)

- Read depth (total and variant-supporting reads)
- Transcript ID
- Clinical significance annotations (ClinVar, OncoKB, etc.)

☐ Germline vs somatic classification
☐ Pathogenicity predictions (SIFT, PolyPhen, CADD scores)

### 3.2 Copy Number Variations (CNVs)

☐ CNV analysis performed (Yes/No)

If Yes, provide:

- Gene-level copy number status (amplification/deletion)
- Copy number value or ratio
- Chromosomal coordinates
- CNV calling method and confidence score

### 3.3 Structural Variants (SVs) & Gene Fusions

☐ Fusion/SV analysis performed (Yes/No)

If Yes, provide:

- Fusion partner genes
- Breakpoint positions
- Supporting read counts
- In-frame vs out-of-frame status
- Known oncogenic fusions (e.g., BCR-ABL, EML4-ALK)

## 4. Additional Molecular Data (if available)

☐ Tumor mutational burden (TMB) score
☐ Microsatellite instability (MSI) status
☐ RNA expression data (if RNA-seq performed)
☐ Gene signature scores (immune, proliferation, etc.)
☐ HLA typing results
☐ Neoantigen prediction data

## 5. Clinical Interpretation & Reporting

☐ Actionable mutations with therapeutic implications
☐ FDA-approved targeted therapies matching variants
☐ Clinical trial eligibility markers
☐ Resistance mutations to standard therapies
☐ Prognostic biomarkers
☐ Variants of unknown significance (VUS) with supporting evidence

## 6. Quality Control Metrics

☐ Total number of reads

- ☐ Percentage of mapped reads
- ☐ Mean/median target coverage
- ☐ Percentage of targets covered at minimum depth (e.g., ≥30x)
- ☐ Tumor purity estimate
- ☐ Contamination assessment
- ☐ Quality control pass/fail status

## 7. Data Format & Delivery Specifications

### Preferred File Formats:

- ☐ VCF (Variant Call Format) for variants
- ☐ BAM/FASTQ files (optional, if raw data needed)
- ☐ JSON format for structured clinical annotations
- ☐ PDF clinical report (human-readable summary)

### Remarks for JSON files:

- Data must be delivered as a single JSON file per sample
- JSON must be UTF-8 encoded
- No free-text fields for genomic coordinates
- Missing mandatory fields will cause automated rejection
- Arrays must be empty [] instead of null