



IT-Unterstützung bei der Sequenzierungsanalyse des Illumina TruSight Oncology 500

Jan Christoph (Bioinformatiker am Lehrstuhl für med. Informatik
bei Prof. Prokosch an der FAU)

15.11.2018

8. Herbsttreffen der Arbeitsgemeinschaft Molekularpathologie der Deutschen
Gesellschaft für Pathologie

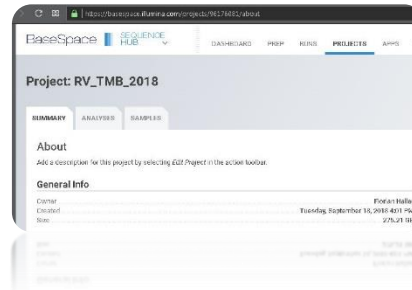
- „The opinions expressed during this presentation are those of the speaker and may not represent the opinions of Illumina. Any uses of Illumina products in this presentation may be uses that have not been cleared or approved by the FDA, or any other applicable regulatory body.“
- „Illumina is reimbursing my travel expenses related to this event.“

Übersicht des Verlaufs



Symbolbild

- s_4.filter
- s_4.bci
- 0218.bcl.bgzf.bci
- 0218.bcl.bgzf
- 0217.bcl.bgzf.bci
- 0217.bcl.bgzf
- 0216.bcl.bgzf.bci
- 0216.bcl.bgzf
- 0215.bcl.bgzf.bci
- 0215.bcl.bgzf
- 0214.bcl.bgzf.bci



```
[TruSight oncology 500, sample Metrics Report]
[HEADER]
SAMPLE_ID RV_TMB_20
REPORT_DATE 2018-10-08
REPORT_TIME 03:06:34L
TSO_500_VERSION 1.0.0.159

[SAMPLE STATUS]
COMPLETED_ALL_STEPS TRUE
FAILED_STEPS NA
STEPS_NOT_EXECUTED NA

[SAMPLE QC METRICS]
PCT_CONTAMINATION_EST 0.0531
CONTAMINATION_P_VALUE 0.9838
CONTAMINATION_SCORE 1543.9338
MEDIAL_EXON_COVERAGE 97
PCT_EXON_SDX 88.263875376302

[TMB METRICS]
CODING_REGION_MB 1.107457
NUMBER_PASSING_ELIGIBLE_NON_SYNONYMOUS_VARIANTS 7
NUMBER_PASSING_ELIGIBLE_VARIANTS 9
TMB 8.126726
NON_SYNONYMOUS_TMB 6.320787
TMB_REGION_MB 1.107457

[MSI METRICS]
TOTAL_MICROSATELLITE_SITES_ASSESSED 26
TOTAL_MICROSATELLITE_SITES_UNSTABLE 2
PERCENTAGE_UNSTABLE_SITES 0.08

[EXPANDED METRICS]
CV_EXON_COVERAGE 0.489197513718311
MEAN_INSERT_SIZE 87
PCT_READ_ENRICHMENT 88.1586299854034
PCT_EXON_150x 9.0514441815325
PCT_CHIMERIC_READS 0.270243841220608
PCT_EXON_100x 47.0237304286101
PCT_EXON_200x 3.248990058539
PCT_ON_TARGET_BASES 47.5123011963193
PCT_ALIGNED_READS 97.9990460328536
TOTAL_PF_READS 11751034
PCT_USABLE_LMI_READS 99.8
MEAN_FAMILY_SIZE 19.49
```

[1] https://upload.wikimedia.org/wikipedia/commons/thumb/8/88/DNA-Sequencers_from_Flickr_57080968.jpg/330px-DNA-Sequencers_from_Flickr_57080968.jpg
 [2] https://de.wikipedia.org/wiki/Datei:CERN_Server.jpg



Sequencer: Illumina TruSight Oncology 500 (TSO500)



- Auf diesem läuft ein Windows Embedded



- Einbinden des Sequencers in das lokale Netzwerk
- Schreiben der Binary-Base-Call(BCL)-Dateien auf Netzlaufwerk (mehrere hundert pro Lauf, nur maschinenlesbar)

```
|  @  '  '  âf@+|  Ö%  
|| AE"¢  n¥!µ¶  @*+--+ <r|<  
êæs  ~{Öx  œie@]  x[Z?c  'Æ  
0h  çoj'm  é-z|'r  +ø?Zw  uv-ù  
c?‡  _-Z#E  âµ>+  Ö||  -  Åo|:  
}iCz  -Æ$>Ä  ·~'É  z[-$i  (E  
ó  úv  @  x]A  
P  ú~+ã  ê  v"è  ]+...i  ±4Hpò  
!i.0t  Aî[>+  $|<  c  d  r  o
```





- Anforderungen (empfohlen / in Erlangen umgesetzt):
 - Hardware:
 - 128 GB Arbeitsspeicher
 - 24 Kerne
 - 1 TB Festplatte
 - Software:
 - RedHat/CentOS Betriebssystem
 - Docker Umgebung
 - Illumina-Docker-Container: *TruSight_Oncology_500_Docker_1.0.0.0.img*
- Kostenpunkt
 - „Miete“ am Rechenzentrum: ca. **1.000 Eur / Jahr**
 - Anschaffung mit Reserve für WES (492 RAM, 48 Kerne, 100TB Festplatte): ca. **75.000 Eur Anschaffung**



- Kopieren der BCL-Dateien des Sequencers auf Server
 - Direkt (ggf. Umweg über Netzablage)
 - Über Illumina-Cloud (BaseSpace Sequence Hub)
- Anpassen einer kleinen Konfigurationsdatei:
 - Adapter
 - Index 1 & Index 2 der Library-Preparation
 - Sample-IDs

```
[Header],,,,,,,,,,
IEMFileVersion,4,,,,,,,,,
Date,19.09.2018,,,,,,,,,
workflow,GenerateFastQ,,,,,,,,,
[Reads],,,,,,,,,,
101,,,,,,,,,
101,,,,,,,,,
[Settings],,,,,,,,,,
Adapter,AGATCGGAAGAGCACACGTCTGAACTCCAGTCA,,,,,,,,,
AdapterRead2,AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT,,,,,,,,,
Read1UMILength,7,,,,,,,,,
Read2UMILength,7,,,,,,,,,
Read1StartFromCycle,9,,,,,,,,,
Read2StartFromCycle,9,,,,,,,,,
[Data]
Lane,Sample_ID,index,index2,Sample_Name,Sample_Plate,Sample
,RV_TMB_08,AAATAATC,GCCTCTAT,,,,,,,,,
,RV_TMB_03,GTAGCTC,CATCCGAA,,,,,,,,,
,RV_TMB_07,ATTCAGAA,AGGCTATA,,,,,,,,,
,RV_TMB_02,GTGAAGCT,TCAGAGCC,,,,,,,,,
,RV_TMB_06,ATGATTAA,GATCTGCT,,,,,,,,,
,RV_TMB_05,ATGCGATAG,ACGAATAA,,,,,,,,,
,RV_TMB_01,ATCCGGAGA,AGGATAGG,,,,,,,,,
,RV_TMB_04,AAATTCGT,TTATGAGT,,,,,,,,,
```



- Starten der Bioinformatik-Pipeline im Docker-Container

```
[root@srv-ap-omics0 TSO500]# ./menu-tso500-1.0.0.0.sh

Please select the runfolder
1) MiniRunFolder          5) RV_TMB_Part3_Harnblase
2) TSO500_Harnblasen     6) TMB_TSO500_Markus_Eckstein
3) RV_TMB_Part1          7) TMB_Agilent_Test
4) TMB_TSO500_Harnblase  8) Quit

please select runfolder:7

Starting docker container using 'TMB_Agilent_Test' as input
2018-11-14 21:27:17 INFO ***** Start TruSight-Oncology-500-pipeline *****
2018-11-14 21:27:17 INFO Workflow500SolidRunner, Version=1.0.0.159, Culture=neutral, PublicKeyToken=null:1.0.0.159
2018-11-14 21:27:17 INFO CommonLogging, Version=1.0.0.11, Culture=neutral, PublicKeyToken=null:1.0.0.11
2018-11-14 21:27:17 INFO Description:Command line args: --analysisFolder /opt/illumina/analysis-folder --workflowCon
fig /opt/illumina/configuration/workflow-config.json --userConfig /opt/illumina/configuration/user-config.json --sampleSheet
/opt/illumina/SampleSheet.csv;
2018-11-14 21:27:17 INFO Description:Start From FASTQ: False;
2018-11-14 21:27:17 DEBUG Description:Creating DAG...;
2018-11-14 21:27:17 DEBUG Description:Checking for cyclical references...;
2018-11-14 21:27:17 DEBUG Description:Executing workflow nodes...;
2018-11-14 21:27:17 DEBUG Description:Checking if node 'ResourceValidation' input parameters are valid...;
2018-11-14 21:27:17 DEBUG Description:Skipping node 'ResourceValidation'.;
2018-11-14 21:27:17 DEBUG Description:Setting node 'ResourceValidation' output parameter validity to 'True'.;
2018-11-14 21:27:17 DEBUG Description:Disabling node 'ResourceValidation'.;
2018-11-14 21:27:17 DEBUG Description:Setting node 'ResourceValidation' output parameter validity to 'False'.;
```

- Im besten Fall: Abholen der Ergebnisse nach ca. 3-5h
 - Ca. 50 GB mitsamt allen möglichen Zwischenformaten
 - Aktuelles „echtes“ Endergebnis: ca. 10KB (0,00001 GB)
- Ansonsten:
 - Fehlersuche: fehlerhafte Konfigurationsdatei, Festplatte voll

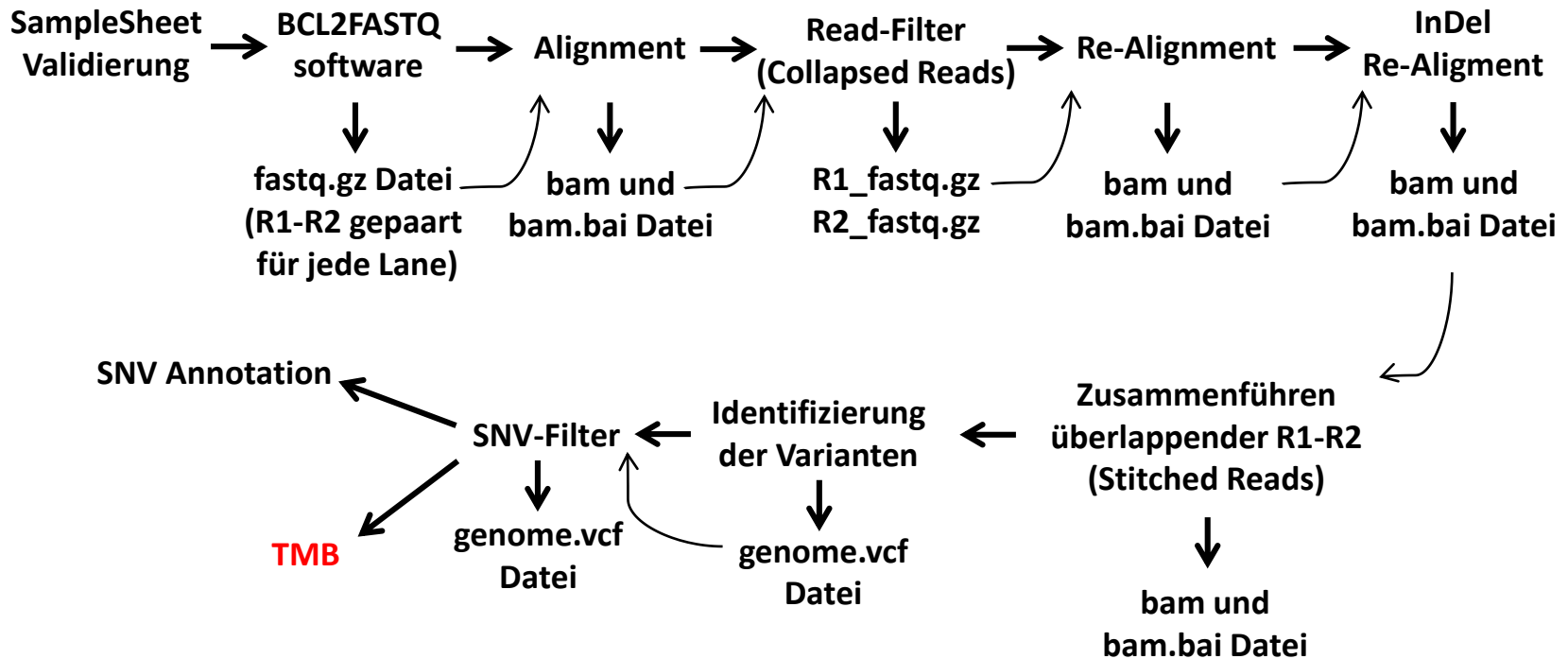


Interner Ablauf der Illumina-Pipeline-Analyse

Auch **open-source-Pipelines** möglich!

<https://github.com/AG-Boerries/MIRACUM-Pipe>

bcl- oder
fastq Datei



Ergebnis (aktuell)

- Textdatei auf Server mit TMB / MSI-Metriken:
- Zwischenformate zur anderweitigen Nutzung: FASTQ
- (offiziell noch keine VCF-files)

```
/home/data/MiniRunFolder/Analysis/Results/0pt2-UMI120-v26-rep1/0pt2-UMI120-v
Trusight oncology 500, sample Metrics Report

[HEADER]
SAMPLE_ID Opt2-UMI120-v26-rep1
REPORT_DATE 2018-08-01
REPORT_TIME 13:30:111
TSO_500_VERSION 1.0.0.159

[SAMPLE STATUS]
COMPLETED_ALL_STEPS True
FAILED_STEPS NA
STEPS_NOT_EXECUTED NA

[SAMPLE QC METRICS]
PCT_CONTAMINATION_EST 0.0131
CONTAMINATION_P_VALUE 1
CONTAMINATION_SCORE 1310.9041
MEDIAN_EXON_COVERAGE 0
PCT_EXON_50X 6.24604544493779

[TMB METRICS]
CODING_REGION_MB 0.084847
NUMBER_PASSING_ELIGIBLE_NON_SYNONOMOUS_VARIANTS 1
NUMBER_PASSING_ELIGIBLE_VARIANTS 1
TMB 11.78592
NON_SYNONOMOUS_TMB 11.78592
TMB_REGION_MB 0.084847

[MSI METRICS]
TOTAL_MICROSATELLITE_SITES_ASSESSED 3
TOTAL_MICROSATELLITE_SITES_UNSTABLE 1
PERCENTAGE_UNSTABLE_SITES 0.33

[EXPANDED METRICS]
CV_EXON_COVERAGE 4.09060737009252
MEDIAN_INSERT_SIZE 162
PCT_READ_ENRICHMENT 57.7978208766967
PCT_EXON_150X 6.14352823078973
PCT_CHIMERIC_READS 0.236772763312398
PCT_EXON_100X 6.1820152171623
PCT_EXON_200X 6.11591949816825
PCT_ON_TARGET_BASES 20.0406790916353
PCT_ALIGNED_READS 96.908987114889
TOTAL_PF_READS 59218582
PCT_USABLE_UMI_READS 99.61
MEAN_FAMILY_SIZE 9.58
```

Zukünftig in Erlangen geplant: Import der VCF-Dateien zur Visualisierung nach cBioPortal



Data Sets Web API R/MATLAB Tutorials FAQ News Visualize Your Data About



Patient: [TCGA-BK-A0CC](#), 69 years old, Endometrial Cancer (Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma), [LIVING](#) (10 months), [DiseaseFree](#) (10 months)

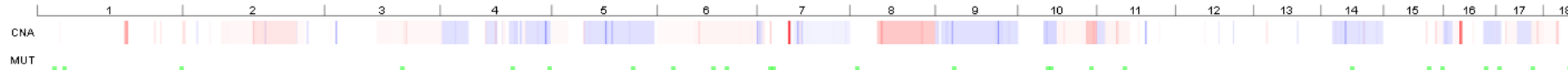
Samples: [TCGA-BK-A0CC-01](#), Stage III

Summary

Clinical Data

Pathology Report

Tissue Image



33 Mutations (page 1 of 4)

Gene	Protein Change	Annotation ▼	Mutation Type	Allele Freq	Copy #	mRNA Expr.
PPP2R1A	S256F		Missense	0.52	Diploid	40%
TP53	Q331*		Nonsense	0.65	ShallowDel	2%
FAT1	E314*		Nonsense	0.64	ShallowDel	47%
EPHA7	H408Q		Missense	0.15	Gain	84%
PHLPP2	R1206Q		Missense	0.62	ShallowDel	85%
DHDDS	A165V		Missense	0.25	Diploid	29%
NSUN4	R10L		Missense	0.44	Diploid	10%
OR2T33	L299Q		Missense	0.13	Diploid	64%
AMOTL2	Q558del		IF del		Gain	99%
FAT4	X3967_splice		Splice	0.31	ShallowDel	53%

Showing 1-10 of 33 Mutations [Show more](#)

380 Copy Number Alterations (page 1 of 38)

Gene	CNA	Annotation ▼	Cytoband	mRNA Expr.
EGFR	AMP		7p11.2	100%
CCNE1	AMP		19q12	88%

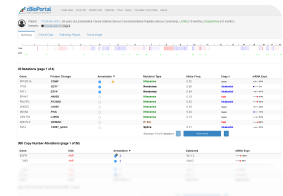
- Entwicklung&Zertifizierung der Pipeline sowie aller sonstigen IT-Komponenten des Workflows als Medizinprodukt?
“For Research Use Only. Not for use in diagnostic procedures.”?!?



- Langzeitarchivierung / Reproduzierbarkeit der Ergebnisse?
 - Welche Daten(formate)?
 - Umgehen bei Updates der Pipeline etc.?



- IT-Unterstützung zur Annotation und Visualisierung der Varianten für die Therapieempfehlung im Molekularen Tumorboard



Entwicklung von Lösungen dafür im Rahmen der Förderinitiative
Medizininformatik => <http://www.medizininformatik-initiative.de>

- speziell im MIRACUM-Konsortium <http://www.miracum.org/>



Danksagung



Medizinisches Zentrum für Informations- und Kommunikationstechnik (MIK)(Erlangen):

- Detlef Kraska
- Terence Schuster
- Roland Oppelt

Lehrstuhl für medizinische Informatik (Erlangen):

- Philipp Unberath
- Ulli Prokosch

Lab for Systems Biology and Systems Medicine (Freiburg)

- Melanie Boerries
- Patrick Metzger

Institut für Pathologie (Erlangen):

- Florian Haller
- Lars Tögel



MITGLIED DER



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

FKZ 01ZZ1801A



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
MEDIZINISCHE FAKULTÄT

Universitätsklinikum
Erlangen

