DATENANALYSE WS22/23

The Story

01.03.2023

Aviral Jain, Jannik Lübke, Leon Geis (WP1)

Daniel Tischler, Noah Leon Stürtz (WP2)

Inhaltsverzeichnis

1. Einleitung
   a. scATACseq
   b. Datenherkunft Gewebe

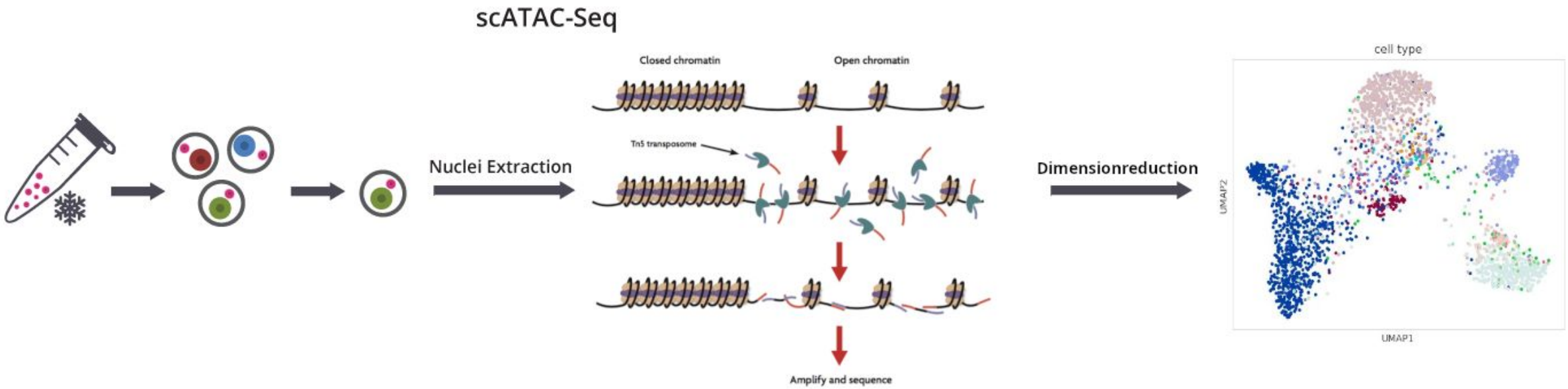
2. WP1
   a. Einlesen der Daten
   b. Visualisierung
      i. Fragment/Violin-Plots
      ii. TSS-Plots
   c. Workflow

3. WP2
   a. Filtern der Referenzdatei
   b. Feature Overlap berechnen
   c. Visualisierung
      i. UMAP
      ii. Scatter/Violin-Plots

4. Einordnung
5. Referenzen

Einleitung

Einleitung in scATACseq



CATlas
Cis-element Atlas

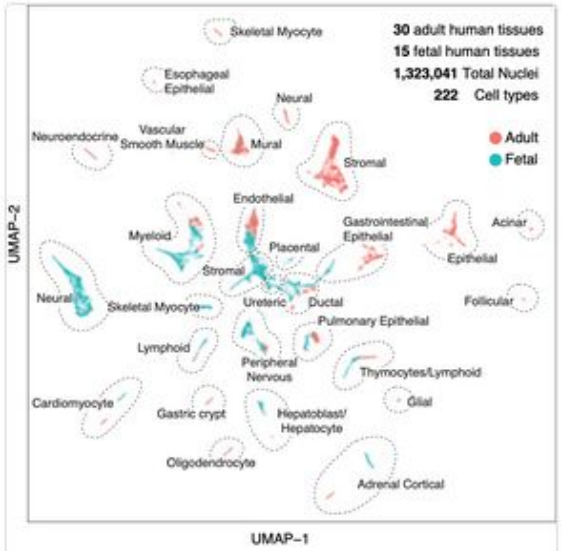
Human Enhancer Atlas

Sample: Human tissues

Method: snATAC-seq

Nuclei count: 1,323,041

Datenherkunft



Cell Clusters

Explore cell clusters in different datasets

Last updated July, 2021

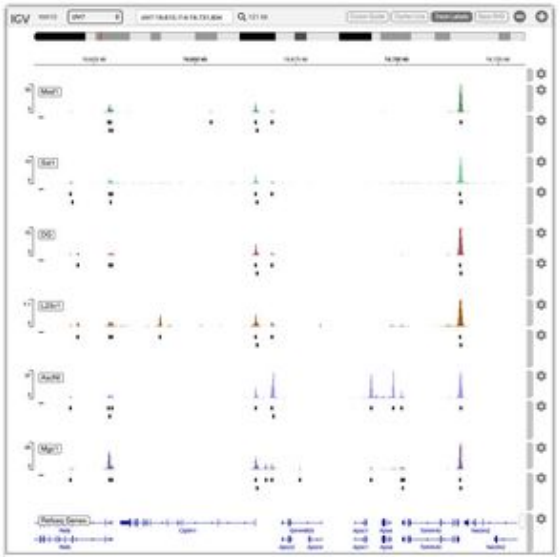
Explore

Class	L3 Cluster	Description
GABAergic	PVGA1	MGE-derived neurogliaform cells, Pvalb
Average TSSE:	20.89	
Average LogUMI:	3.63	
More Info		
GABAergic	SSTGA5	MGE-derived neurogliaform cells, Sst
GABAergic	LAMGA2	CGE-derived neurogliaform cells, Lamp5
GABAergic	PVGA6	MGE-derived neurogliaform cells, Pvalb
GABAergic	VIPGA3	CGE-derived neurogliaform cells, Vip
GABAergic	LAMGA4	CGE-derived neurogliaform cells, Lamp5
GABAergic	SSTGA1	MGE-derived neurogliaform cells, Sst
GABAergic	VIPGA1	Cholinergic neurons, Cortex
GABAergic	SSTGA4	MGE-derived neurogliaform cells, Sst
GABAergic	SSTGA3	MGE-derived neurogliaform cells, Sst

Cell Types

Explore detailed information for every cell type

Explore



Explore Tracks

Load signal tracks and compare between cell types/regions

Explore

http://catlas.org/catlas_humanenhancer/#!/ (Aufgerufen: 25.02.23)

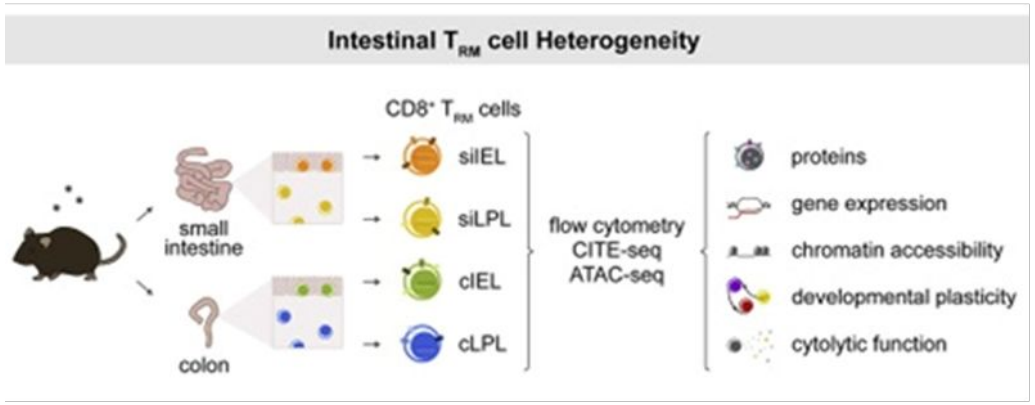
RESOURCE | VOLUME 56, ISSUE 1, P207-223.E8, JANUARY 10, 2023

Small intestine and colon tissue-resident memory CD8$^+$ T cells exhibit molecular heterogeneity and differential dependence on Eomes

Yun Hsuan Lin [6] • Han G. Duong [6] • Abigail E. Limary [6] • ... Gene W. Yeo • Ananda W. Goldrath • John T. Chang [7] • Show all authors • Show footnotes

Highlights

- Small intestine (SI) and colon CD8$^+$ T_{RM} cells are molecularly and functionally distinct
- Anatomically distinct SI T_{RM} cells exhibit disparate degrees of developmental plasticity
- Eomes supports maintenance of established T_{RM} cells in the SI, but not in the colon



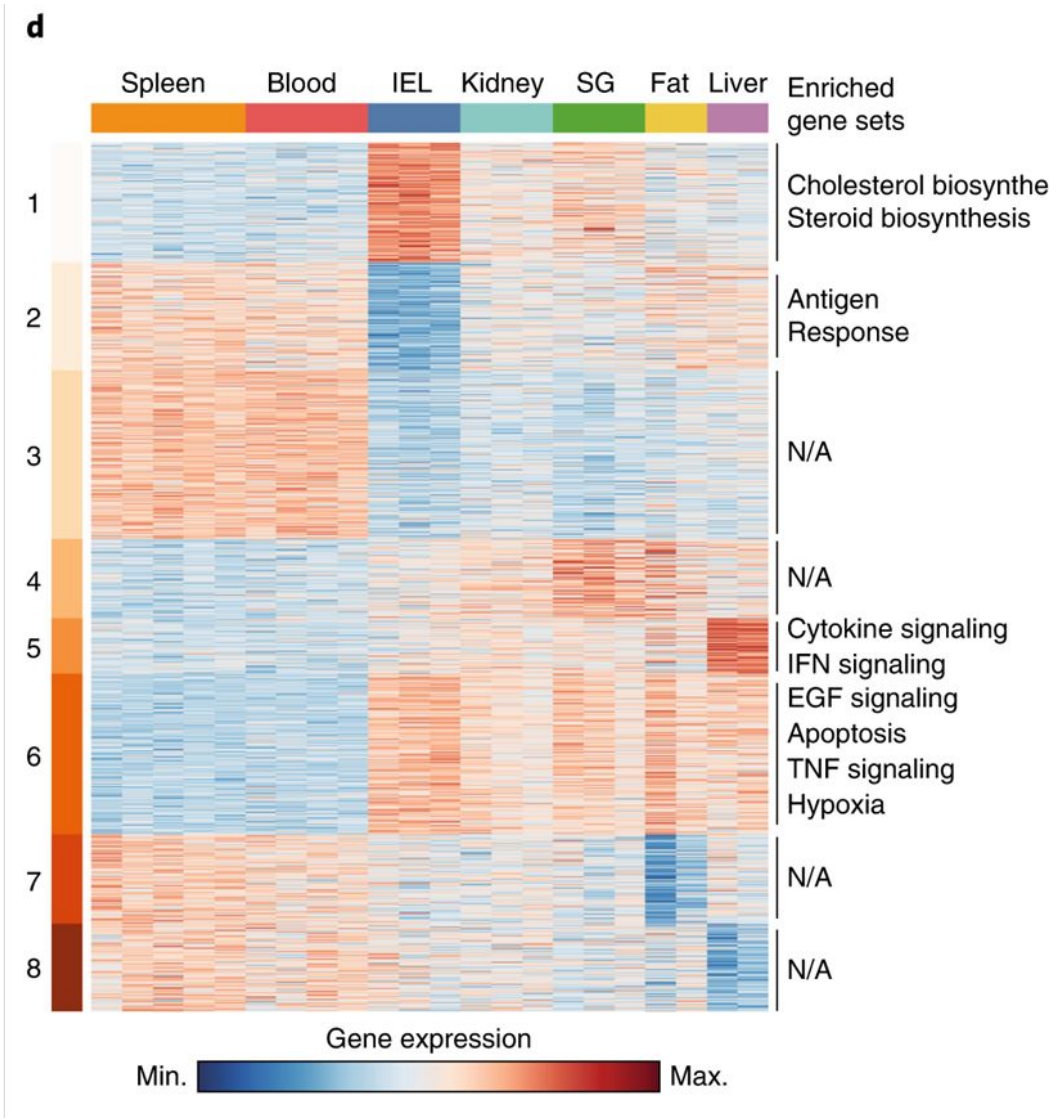
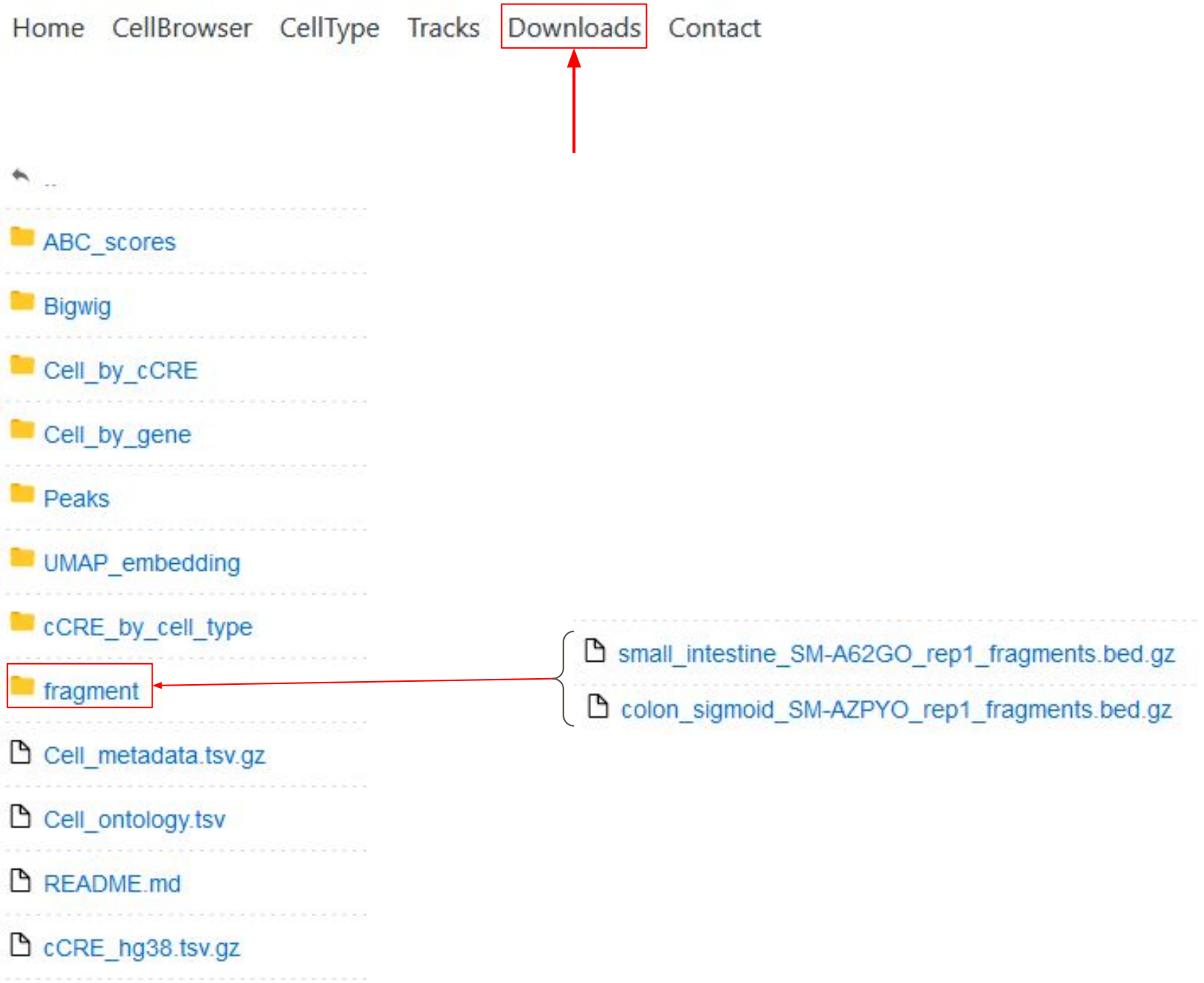
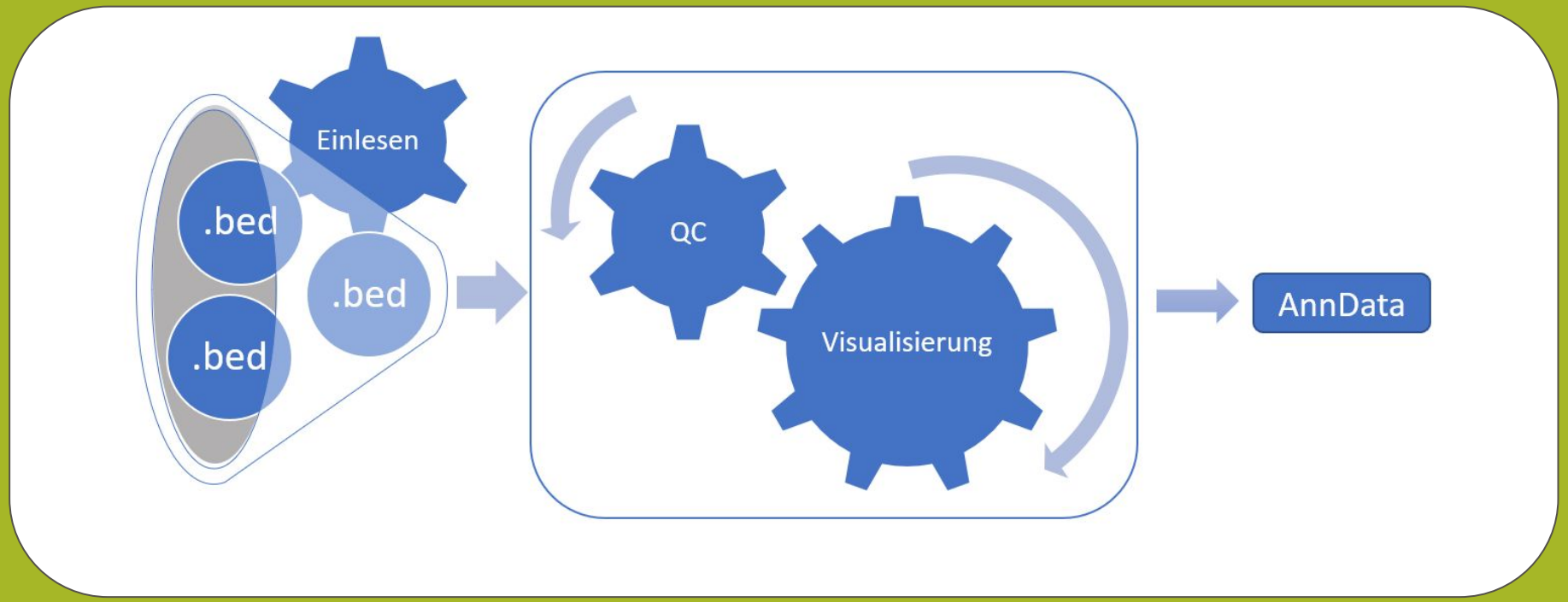


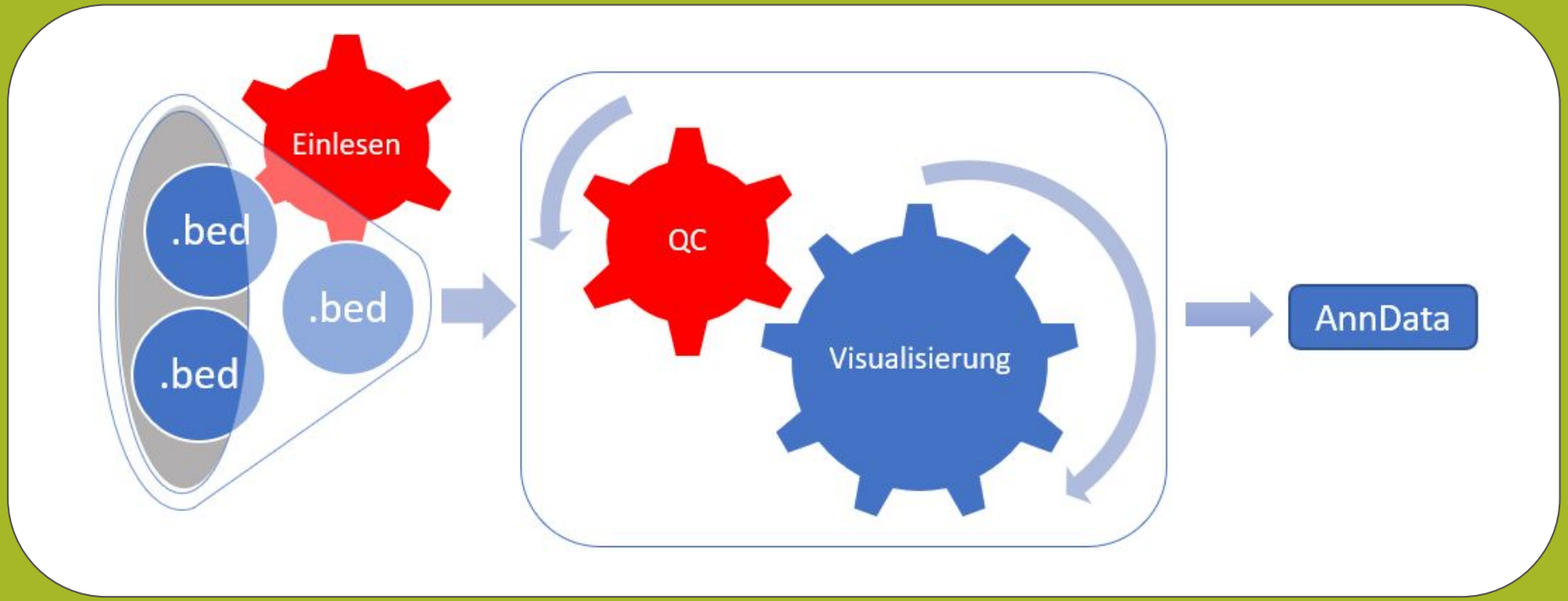
image: https://www.nature.com/articles/s41590-022-01229-8

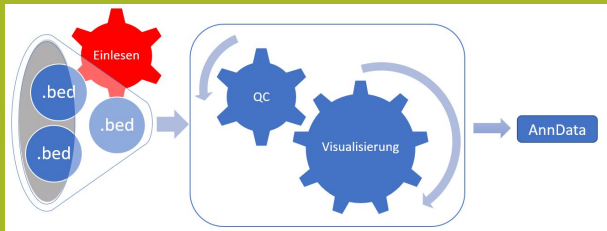

Home
CellBrowser
CellType
Tracks
Downloads
Contact
..
ABC_scores
Bigwig
Cell_by_cCRE
Cell_by_gene
Peaks
UMAP_embedding
cCRE_by_cell_type
fragment
Cell_metadata.tsv.gz
Cell_ontology.tsv
README.md
cCRE_hg38.tsv.gz
small_intestine_SM-A62GO_rep1_fragments.bed.gz
colon_sigmoid_SM-AZPYO_rep1_fragments.bed.gz

WP1



Einlesen der Daten





Einlesen der Daten

small_intestine_SM-A62GO_rep1_fragments.bed.gz

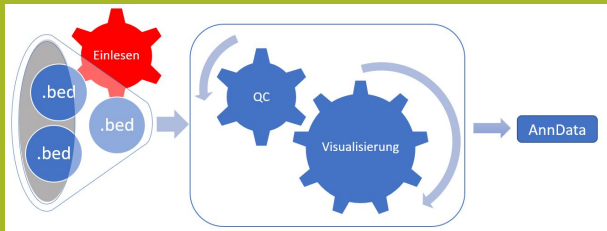
Entpackt

chr1	6935282	6935441	AAACGCAAGCAAACCCGAGATA	1	.
chr1	9726090	9726387	AAACGCAAGCAAACCCGAGATA	3	.
chr1	10399318	10399366	AAACGCAAGCAAACCCGAGATA	5	.
chr1	16036235	16036706	AAACGCAAGCAAACCCGAGATA	2	.
chr1	18287580	18287661	AAACGCAAGCAAACCCGAGATA	5	.
chr1	27933424	27933478	AAACGCAAGCAAACCCGAGATA	1	.
chr1	31782619	31782662	AAACGCAAGCAAACCCGAGATA	1	.
chr1	42381021	42381055	AAACGCAAGCAAACCCGAGATA	5	.
chr1	44422250	44422359	AAACGCAAGCAAACCCGAGATA	1	.
chr1	44674226	44674468	AAACGCAAGCAAACCCGAGATA	2	.
chr1	51497877	51497941	AAACGCAAGCAAACCCGAGATA	5	.
chr1	51700660	51700850	AAACGCAAGCAAACCCGAGATA	1	.
chr1	65638758	65638995	AAACGCAAGCAAACCCGAGATA	1	.
chr1	67833091	67833471	AAACGCAAGCAAACCCGAGATA	2	.
chr1	75550605	75550843	AAACGCAAGCAAACCCGAGATA	3	.
chr1	78249177	78249223	AAACGCAAGCAAACCCGAGATA	2	.
chr1	78352912	78353158	AAACGCAAGCAAACCCGAGATA	3	.
chr1	78621974	78622102	AAACGCAAGCAAACCCGAGATA	1	.
chr1	79121107	79121151	AAACGCAAGCAAACCCGAGATA	3	.
chr1	79429471	79429860	AAACGCAAGCAAACCCGAGATA	1	.
chr1	85179609	85179775	AAACGCAAGCAAACCCGAGATA	3	.
chr1	85329067	85329271	AAACGCAAGCAAACCCGAGATA	5	.
chr1	85333390	85333845	AAACGCAAGCAAACCCGAGATA	1	.
chr1	91352688	91352901	AAACGCAAGCAAACCCGAGATA	3	.
chr1	91783703	91783991	AAACGCAAGCAAACCCGAGATA	4	.
chr1	92204277	92204334	AAACGCAAGCAAACCCGAGATA	7	.
chr1	96213696	96213866	AAACGCAAGCAAACCCGAGATA	4	.
chr1	99685833	99686142	AAACGCAAGCAAACCCGAGATA	2	.
chr1	159700498	159700854	AAACGCAAGCAAACCCGAGATA	1	.
chr1	175481965	175482011	AAACGCAAGCAAACCCGAGATA	3	.

colon_sigmoid_SM-AZPYO_rep1_fragments.bed.gz

Entpackt

chr1	634016	634074	AACGAGAGCTAAACCCGAGATA	5	.
chr1	941452	941626	AACGAGAGCTAAACCCGAGATA	3	.
chr1	1307954	1308516	AACGAGAGCTAAACCCGAGATA	3	.
chr1	1324856	1324910	AACGAGAGCTAAACCCGAGATA	3	.
chr1	1399484	1399513	AACGAGAGCTAAACCCGAGATA	1	.
chr1	1906117	1906226	AACGAGAGCTAAACCCGAGATA	3	.
chr1	1919304	1919371	AACGAGAGCTAAACCCGAGATA	2	.
chr1	2227624	2227735	AACGAGAGCTAAACCCGAGATA	1	.
chr1	2548849	2548969	AACGAGAGCTAAACCCGAGATA	3	.
chr1	3258775	3258833	AACGAGAGCTAAACCCGAGATA	1	.
chr1	3258850	3259030	AACGAGAGCTAAACCCGAGATA	1	.
chr1	3553030	3553137	AACGAGAGCTAAACCCGAGATA	1	.
chr1	4999060	4999215	AACGAGAGCTAAACCCGAGATA	1	.
chr1	5685796	5685884	AACGAGAGCTAAACCCGAGATA	2	.
chr1	6262136	6262276	AACGAGAGCTAAACCCGAGATA	1	.
chr1	6358655	6358720	AACGAGAGCTAAACCCGAGATA	4	.
chr1	6785106	6785243	AACGAGAGCTAAACCCGAGATA	3	.
chr1	7465559	7465656	AACGAGAGCTAAACCCGAGATA	4	.
chr1	7477161	7477199	AACGAGAGCTAAACCCGAGATA	1	.
chr1	8426687	8427106	AACGAGAGCTAAACCCGAGATA	1	.
chr1	8727180	8727380	AACGAGAGCTAAACCCGAGATA	2	.
chr1	9405318	9405369	AACGAGAGCTAAACCCGAGATA	2	.
chr1	10430540	10430726	AACGAGAGCTAAACCCGAGATA	4	.
chr1	11315115	11315418	AACGAGAGCTAAACCCGAGATA	1	.
chr1	11786211	11786270	AACGAGAGCTAAACCCGAGATA	2	.
chr1	11787182	11787244	AACGAGAGCTAAACCCGAGATA	4	.
chr1	14511814	14511962	AACGAGAGCTAAACCCGAGATA	4	.
chr1	14730835	14730974	AACGAGAGCTAAACCCGAGATA	1	.
chr1	15691254	15691710	AACGAGAGCTAAACCCGAGATA	1	.
chr1	16009875	16010177	AACGAGAGCTAAACCCGAGATA	1	.



Einlesen der Daten

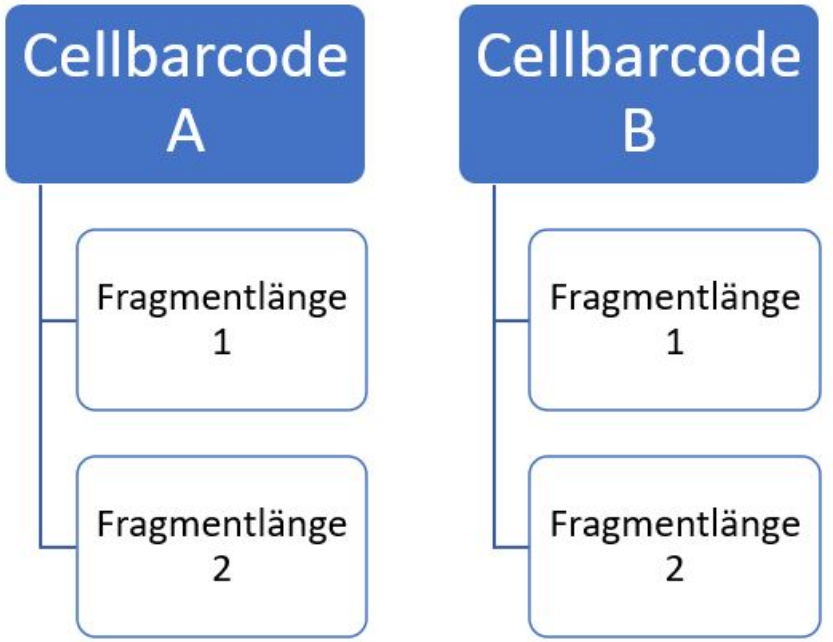
small_intestine_SM-A62GO_rep1_fragments.bed.gz

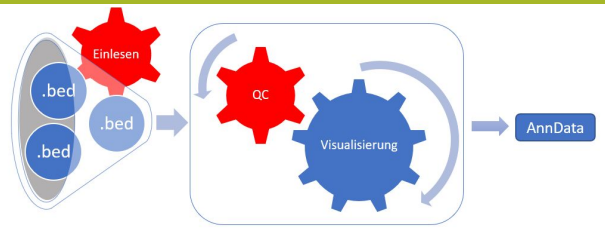
chr1	6935282	6935441	AAACGCAAGCAAACCCGAGATA	1	.
chr1	9726090	9726387	AAACGCAAGCAAACCCGAGATA	3	.
chr1	10399318	10399366	AAACGCAAGCAAACCCGAGATA	5	.
chr1	16036235	16036706	AAACGCAAGCAAACCCGAGATA	2	.
chr1	18287580	18287661	AAACGCAAGCAAACCCGAGATA	5	.
chr1	27933424	27933478	AAACGCAAGCAAACCCGAGATA	1	.
chr1	31782619	31782662	AAACGCAAGCAAACCCGAGATA	1	.
chr1	42381021	42381055	AAACGCAAGCAAACCCGAGATA	5	.
chr1	44422250	44422359	AAACGCAAGCAAACCCGAGATA	1	.
chr1	44674226	44674468	AAACGCAAGCAAACCCGAGATA	2	.
chr1	51497877	51497941	AAACGCAAGCAAACCCGAGATA	5	.
chr1	51700660	51700850	AAACGCAAGCAAACCCGAGATA	1	.
chr1	65638758	65638995	AAACGCAAGCAAACCCGAGATA	1	.
chr1	67833091	67833471	AAACGCAAGCAAACCCGAGATA	2	.
chr1	75550605	75550843	AAACGCAAGCAAACCCGAGATA	3	.
chr1	78249177	78249223	AAACGCAAGCAAACCCGAGATA	2	.
chr1	78352912	78353158	AAACGCAAGCAAACCCGAGATA	3	.
chr1	78621974	78622102	AAACGCAAGCAAACCCGAGATA	1	.
chr1	79121107	79121151	AAACGCAAGCAAACCCGAGATA	3	.
chr1	79429471	79429860	AAACGCAAGCAAACCCGAGATA	1	.
chr1	85179609	85179775	AAACGCAAGCAAACCCGAGATA	3	.
chr1	85329067	85329271	AAACGCAAGCAAACCCGAGATA	5	.
chr1	85333390	85333845	AAACGCAAGCAAACCCGAGATA	1	.
chr1	91352688	91352901	AAACGCAAGCAAACCCGAGATA	3	.
chr1	91783703	91783991	AAACGCAAGCAAACCCGAGATA	4	.
chr1	92204277	92204334	AAACGCAAGCAAACCCGAGATA	7	.
chr1	96213696	96213866	AAACGCAAGCAAACCCGAGATA	4	.
chr1	99685833	99686142	AAACGCAAGCAAACCCGAGATA	2	.
chr1	159700498	159700854	AAACGCAAGCAAACCCGAGATA	1	.
chr1	175481965	175482011	AAACGCAAGCAAACCCGAGATA	3	.

Fragmentstart

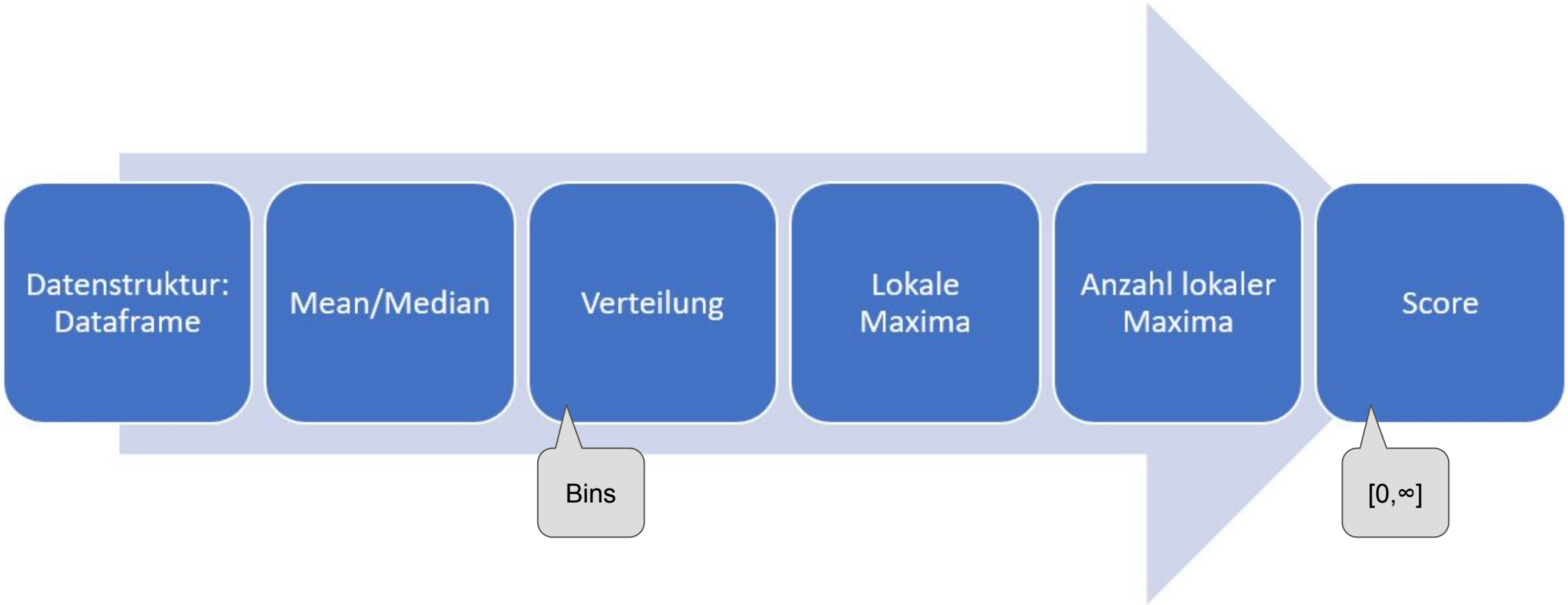
Fragmentstop

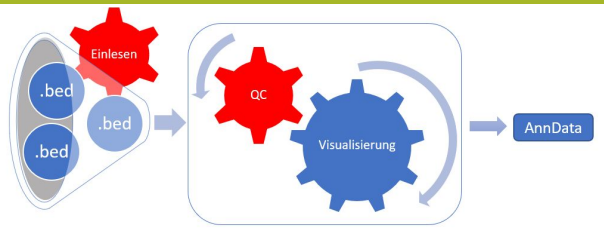
Cellbarcode





Einlesen, ein großer Prozess

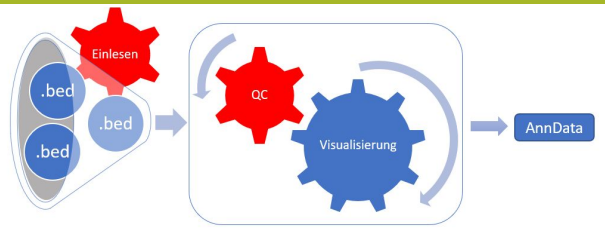




Einlesen, ein großer Prozess

small_intestine_SM-A62GO_rep1_fragments.bed

	Mean	Median	Fragments	Fragment-Count	Distribution	Maxima	Maxima-Count	Score
AAACGCAAGCAAACCCGAGATA	160.44	127.0	[159, 297, 48, 471, 81, 54, 43, 34, 109,...	448	[39, 66, 64, 37, 25, 18, 25, 28, 26, 30, ...	[1, 9, 14, 19]	4	6.31143
AAACGCAAGCAAACCTAAGTGG	148.42	104.0	[34, 87, 535, 75, 152, 53, 184, 109, 43,...	12	[2, 1, 2, 2, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, ...	[23]	1	216.74858
AAACGCAAGCAAACGGATCAGT	114.0	98.5	[71, 195, 181, 40, 110, 87]	6	[1, 0, 2, 1, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, ...	[2, 7]	2	135.62088
AAACGCAAGCAAACGTCCCGTT	90.0	82.5	[83, 82, 92, 174, 45, 64]	6	[1, 1, 2, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, ...	[2, 6]	2	148.01094
AAACGCAAGCAAACTAGCCCTA	100.58	78.5	[126, 259, 220, 74, 51, 121, 54, 85, 38,...	12	[2, 3, 3, 0, 2, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, ...	[]	0	inf
AAACGCAAGCAAAGCACTAGCG	122.17	69.0	[38, 46, 273, 75, 63, 238]	6	[1, 2, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, ...	[1]	1	185.18111
AAACGCAAGCAAAGCCCACGAC	134.6	81	[202, 55, 39, 36, 81, 45, 81, 68, 371, 1...	15	[3, 2, 3, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1...	[15]	1	132.19844
AAACGCAAGCAAAGCGGGAGCT	74.42	70.5	[42, 98, 63, 42, 64, 43, 95, 93, 115, 10...	12	[3, 3, 1, 4, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[3]	1	113.24369
AAACGCAAGCAAAGGAACAGAC	102.73	66	[68, 479, 66, 52, 86, 65, 100, 65, 71, 3...	11	[2, 4, 3, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[1, 20]	2	187.99820
AAACGCAAGCAAAGGATCGGCT	154.6	118.0	[42, 177, 43, 194, 441, 30, 235, 33, 33,...	2042	[257, 308, 244, 169, 141, 117, 121, 97, 1...	[1, 16, 23, 26]	4	11.85171
AAACGCAAGCAAAGGGATGCCA	63.5	63.5	[56, 71]	2	[0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[]	0	inf
AAACGCAAGCAAAGTCGCGTGT	97.93	59	[46, 39, 125, 68, 136, 45, 43, 59, 99, 2...	15	[5, 3, 2, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, ...	[0, 9, 16]	3	76.51261
AAACGCAAGCAAATAAGGCCAG	153.14	121.0	[110, 98, 247, 354, 123, 249, 291, 131, ...	14	[2, 1, 1, 2, 3, 0, 0, 0, 0, 1, 2, 0, 1, 0, 1, ...	[0, 4, 10]	3	89.04695
AAACGCAAGCAAATGCTACGGG	226.5	190.0	[220, 145, 321, 445, 68, 160]	6	[0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, ...	[2, 13, 19]	3	139.24860
AAACGCAAGCAAATTGAGGAGG	145.38	85.5	[76, 455, 32, 28, 251, 95, 162, 64]	8	[2, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, ...	[0, 6, 10, 19]	4	114.42014
AAACGCAAGCAACAGGCGGGTA	103.64	69.0	[420, 40, 82, 130, 48, 64, 198, 40, 87, ...	14	[4, 3, 3, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, ...	[0, 17]	2	152.85761
AAACGCAAGCAACATGAAGCGC	161.46	142	[226, 48, 88, 278, 179, 69, 57, 73, 122,...	13	[0, 2, 3, 0, 1, 1, 0, 1, 0, 2, 1, 1, 0, 0, 1, 0...	[2, 9, 14]	3	83.35447
AAACGCAAGCAACCACCTAAAG	159.44	114	[46, 199, 54, 196, 38, 56, 35, 412, 336...	17299	[2129, 2877, 2015, 1486, 1036, 881, 82...	[1]	1	14.12116
AAACGCAAGCAACCATGCATGA	137.6	105	[105, 383, 52, 38, 110]	5	[1, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[3, 16]	2	201.06398
AAACGCAAGCAACCCACTATCT	159.76	116.0	[183, 110, 77, 43, 35, 355, 64, 599, 25...	2742	[301, 429, 352, 246, 151, 128, 172, 160...	[1, 6]	2	6.18965

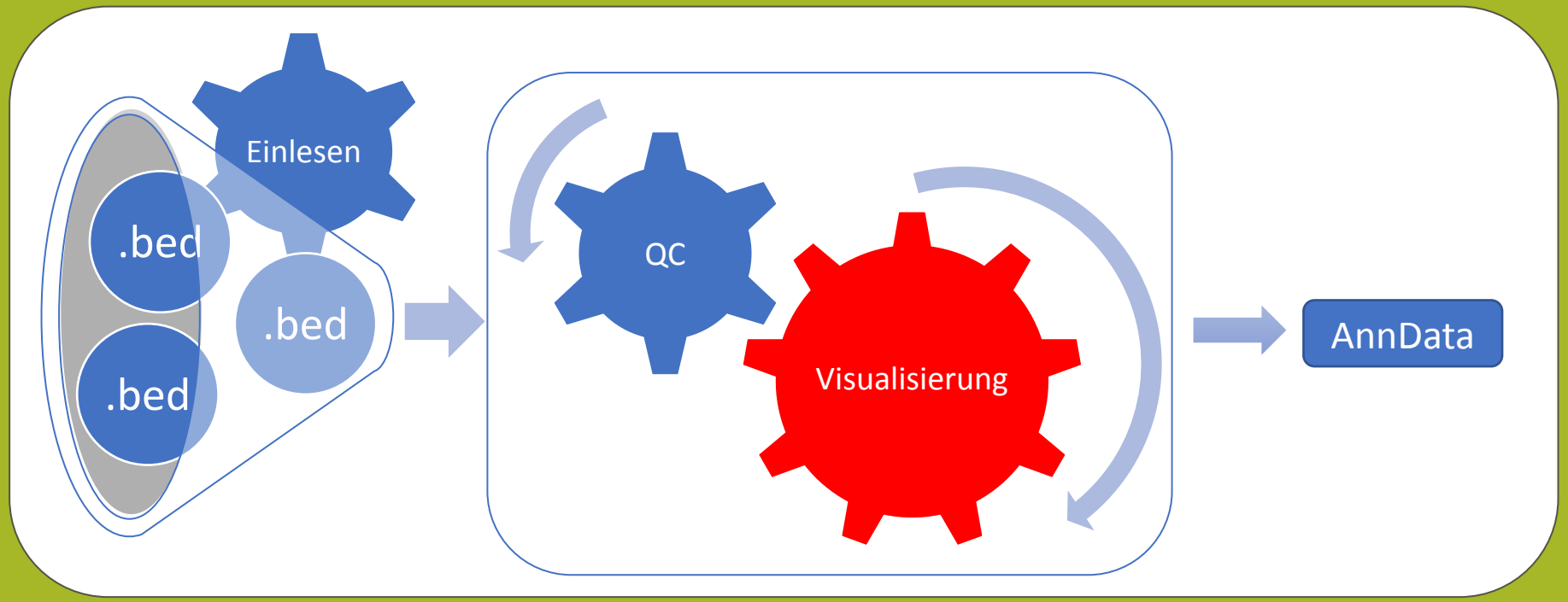


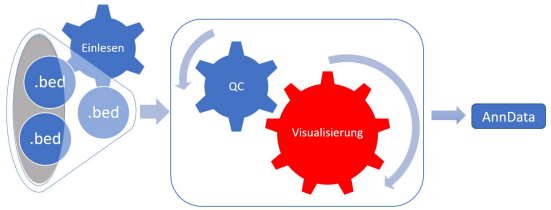
Einlesen, ein großer Prozess

colon_sigmoid_SM-AZPYO_rep1_fragments.bed

	Mean	Median	Fragments	Fragment-Count	Distribution	Maxima	Maxima-Count	Score
AACGAGAGCTAAACCCGAGATA	163.97	118.0	[58, 174, 562, 54, 29, 109, 67, 111, 120...	2450	[308, 449, 306, 194, 160, 137, 115, 132...	[1, 14]	2	21.25491
AACGAGAGCTAAACCTAAGTGG	119.88	59.5	[71, 438, 35, 77, 48, 224, 34, 32]	8	[4, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, ...	[0, 8, 16]	3	111.82329
AACGAGAGCTAAACGGATCAGT	147.0	80.5	[46, 42, 344, 74, 87, 375, 151, 57]	8	[2, 1, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, ...	[5]	1	109.00523
AACGAGAGCTAAACGTCCCGTT	77.5	61.5	[30, 117, 171, 86, 53, 54, 29, 135, 31, 69]	10	[3, 3, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[]	0	inf
AACGAGAGCTAAACTAGCCCTA	164.05	115	[33, 87, 78, 70, 52, 65, 82, 116, 44, 12...	2729	[453, 474, 315, 188, 138, 144, 128, 134...	[1, 12]	2	14.62772
AACGAGAGCTAAAGCACTAGCG	169.62	129.0	[84, 169, 73, 242, 254, 170, 73, 259, 1...	3496	[466, 532, 400, 302, 224, 211, 182, 18...	[1, 14]	2	20.32876
AACGAGAGCTAAAGCCCACGAC	111.5	71.0	[52, 144, 232, 44, 288, 152, 42, 59, 62...	12	[3, 3, 1, 0, 2, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, ...	[4]	1	99.69389
AACGAGAGCTAAAGCGGGAGCT	155.67	61	[351, 55, 61]	3	[0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,...	[1, 13]	2	307.47881
AACGAGAGCTAAAGGAACAGAC	127.91	64	[26, 164, 463, 154, 64, 47, 235, 48, 10...	11	[4, 2, 0, 1, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, ...	[0, 5, 8, 17]	4	104.54585
AACGAGAGCTAAAGGATCGGCT	105.4	74	[126, 218, 74, 37, 72]	5	[1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ...	[7]	1	130.77236
AACGAGAGCTAAAGGGATGCCA	73.0	73.0	[63, 83]	2	[0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[]	0	inf
AACGAGAGCTAAAGTCGCGTGT	88.36	78	[34, 131, 46, 121, 76, 78, 149, 58, 107, ...	11	[2, 2, 2, 3, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[3]	1	114.18347
AACGAGAGCTAAATAAGGCCAG	164.71	123	[155, 43, 350, 363, 247, 38, 174, 77, 9...	4233	[471, 723, 547, 379, 288, 264, 241, 22...	[1]	1	16.15799
AACGAGAGCTAAATGCTACGGG	94.25	94.0	[48, 51, 137, 141]	4	[1, 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[4]	1	184.47020
AACGAGAGCTAAATTGAGGAGG	140.7	86.0	[354, 292, 174, 42, 39, 70, 57, 102, 23...	10	[3, 2, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, ...	[0, 13]	2	154.55238
AACGAGAGCTAACAGGCGGGTA	171.5	80.0	[492, 34, 74, 86]	4	[1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...	[2, 18]	2	315.27936
AACGAGAGCTAACATGAAGCGC	120.1	77.5	[73, 60, 252, 82, 59, 49, 212, 67, 143, ...	10	[0, 5, 1, 0, 1, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, ...	[1, 7]	2	86.68518
AACGAGAGCTAACCACCTAAAG	173.99	128.0	[50, 274, 66, 156, 41, 164, 39, 457, 62,...	3840	[516, 592, 456, 317, 242, 208, 170, 18...	[1, 20]	2	38.32077
AACGAGAGCTAACCATGCATGA	116.38	60.0	[219, 32, 204, 73, 274, 47, 37, 45]	8	[4, 1, 0, 0, 0, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, ...	[0, 7, 10]	3	116.43992
AACGAGAGCTAACCCACTATCT	138.4	85.0	[93, 598, 81, 53, 35, 59, 212, 64, 89, 1...	10	[1, 3, 3, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ...	[7, 22]	2	176.32356

Visualisierung

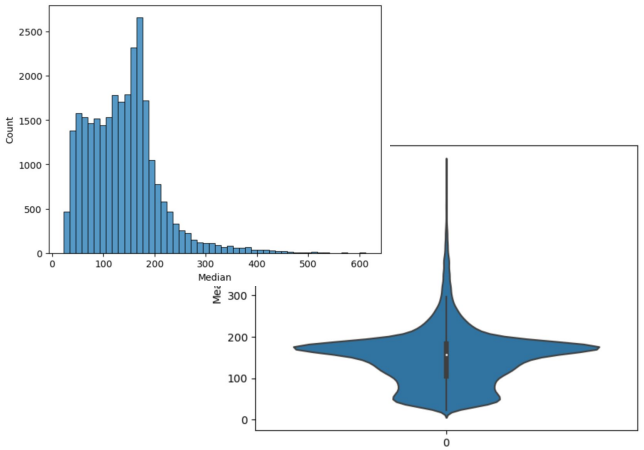




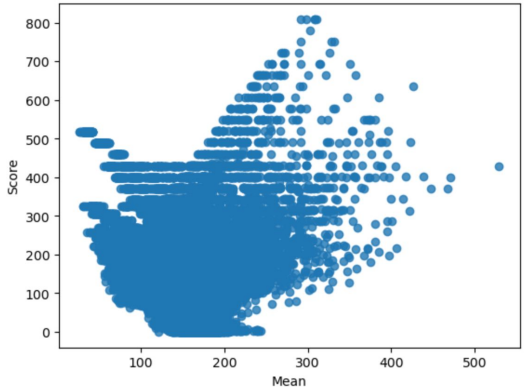
Visualisierungsformen

Verteilung der Quality-Control-Parameter

- alleinstehend
  - Histogramme
  - Violinen-Diagramme

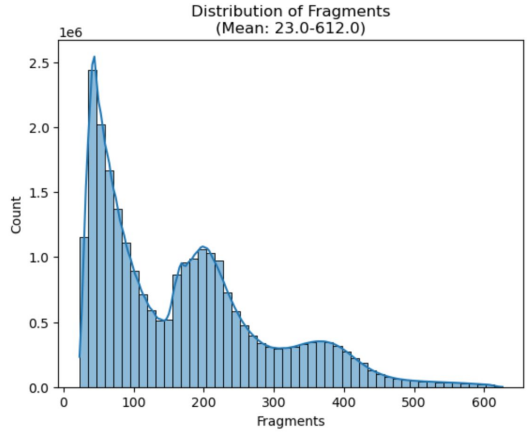


- vergleichend
  - Streu-Diagramme

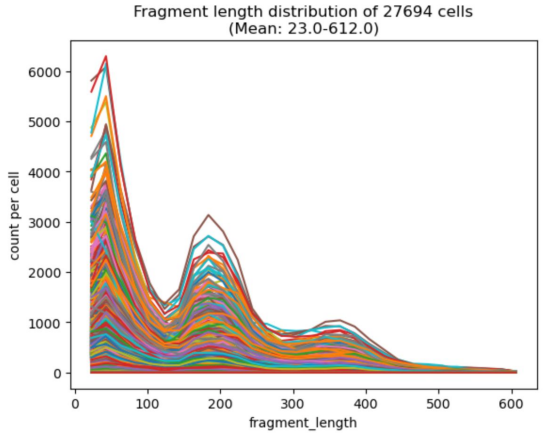


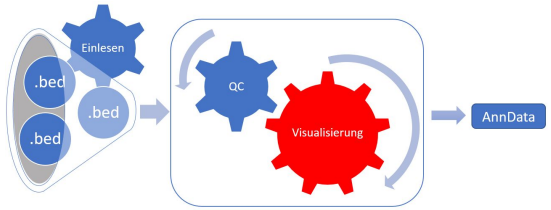
Verteilung der Fragment-Längen

- ohne Zell-Identität als Histogram



- mit Zell-Identität als gruppierte Einzel-Plots

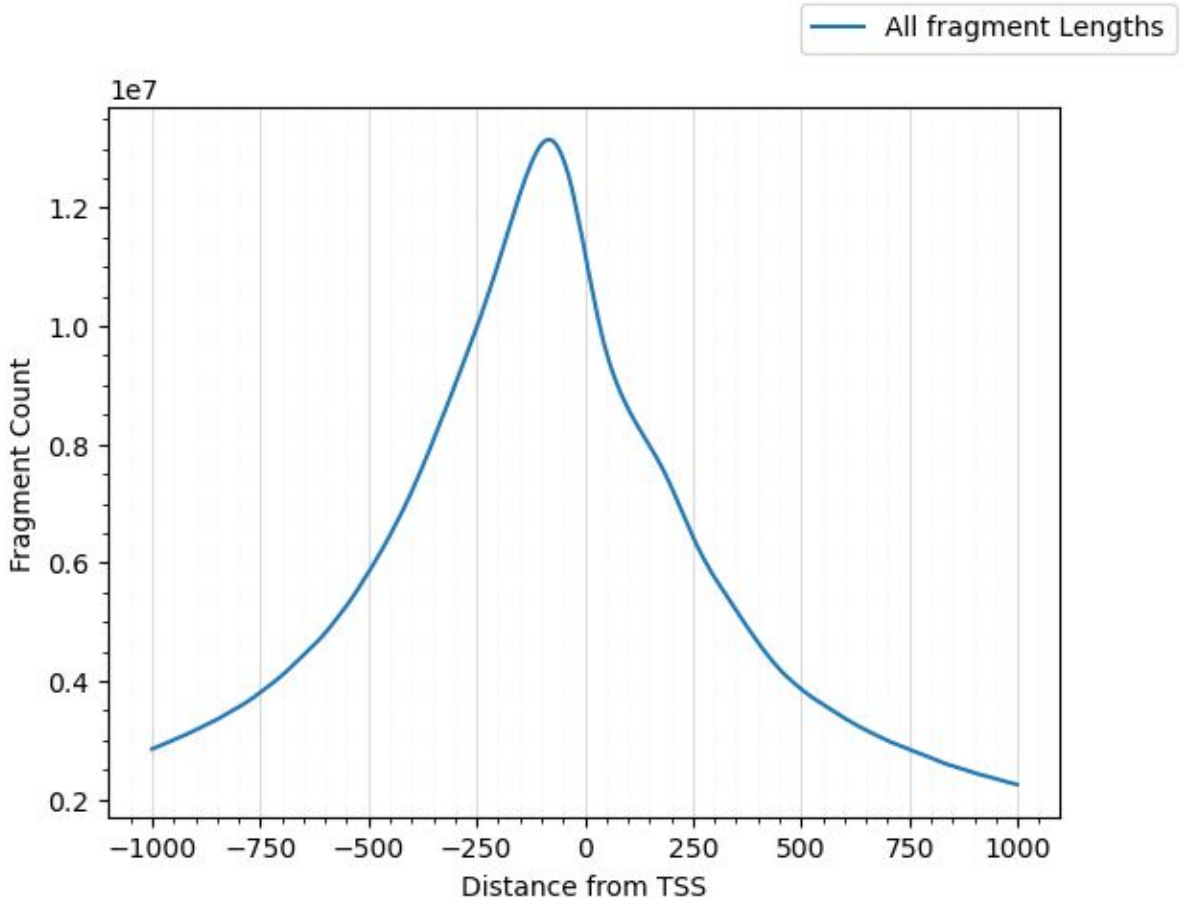




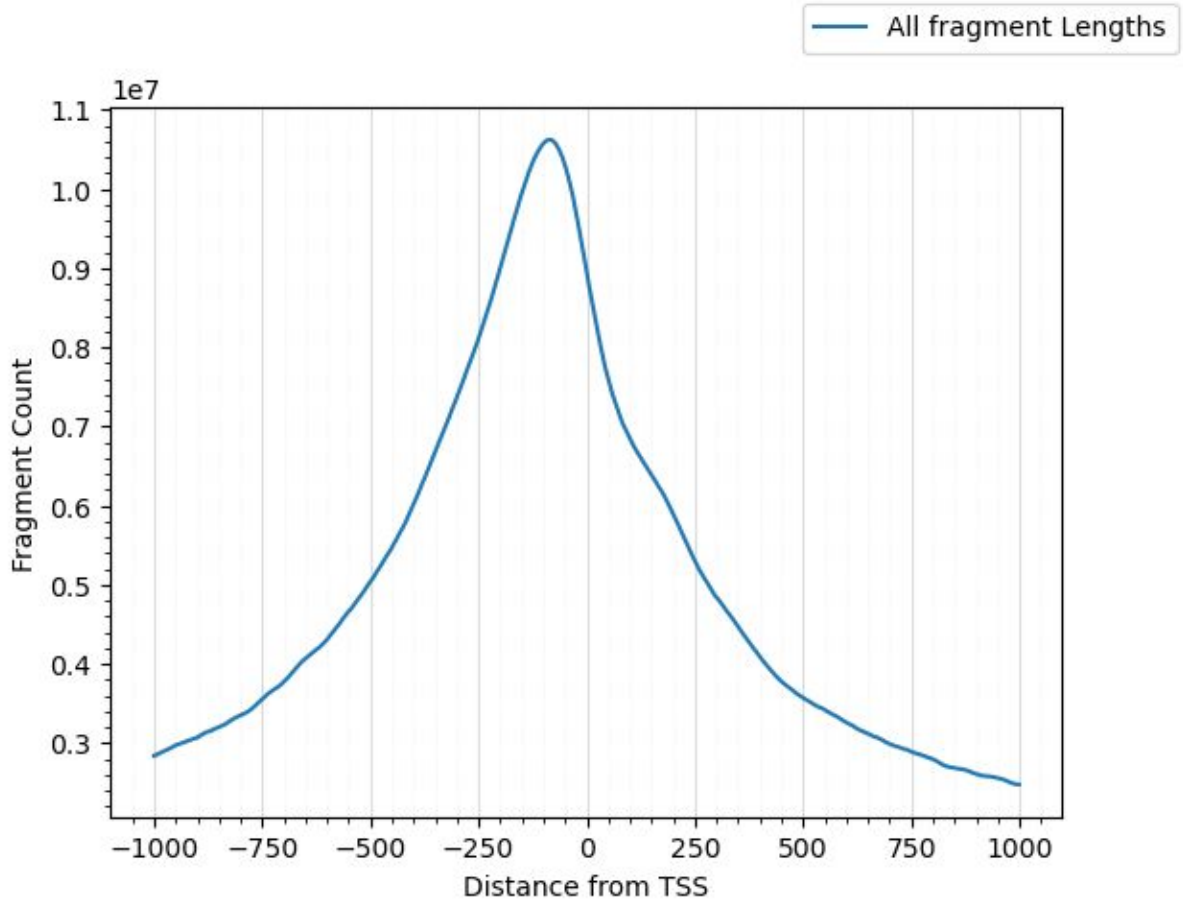
TSS Profile Plots

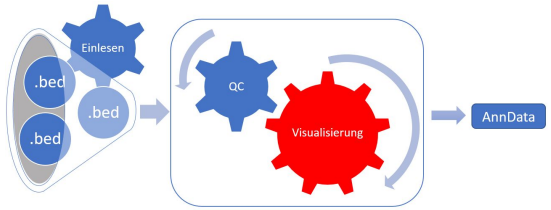
Alle Fragmente

Small Intestine



Colon

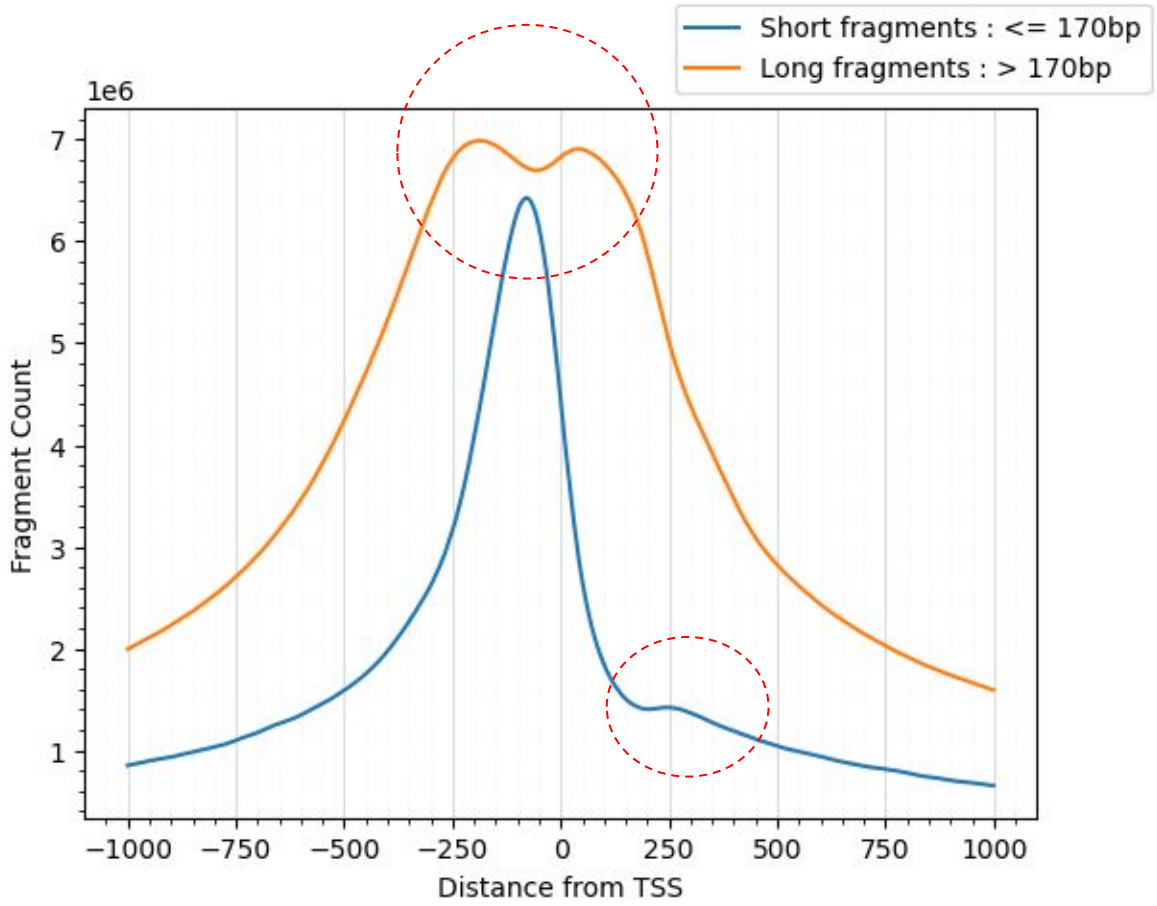




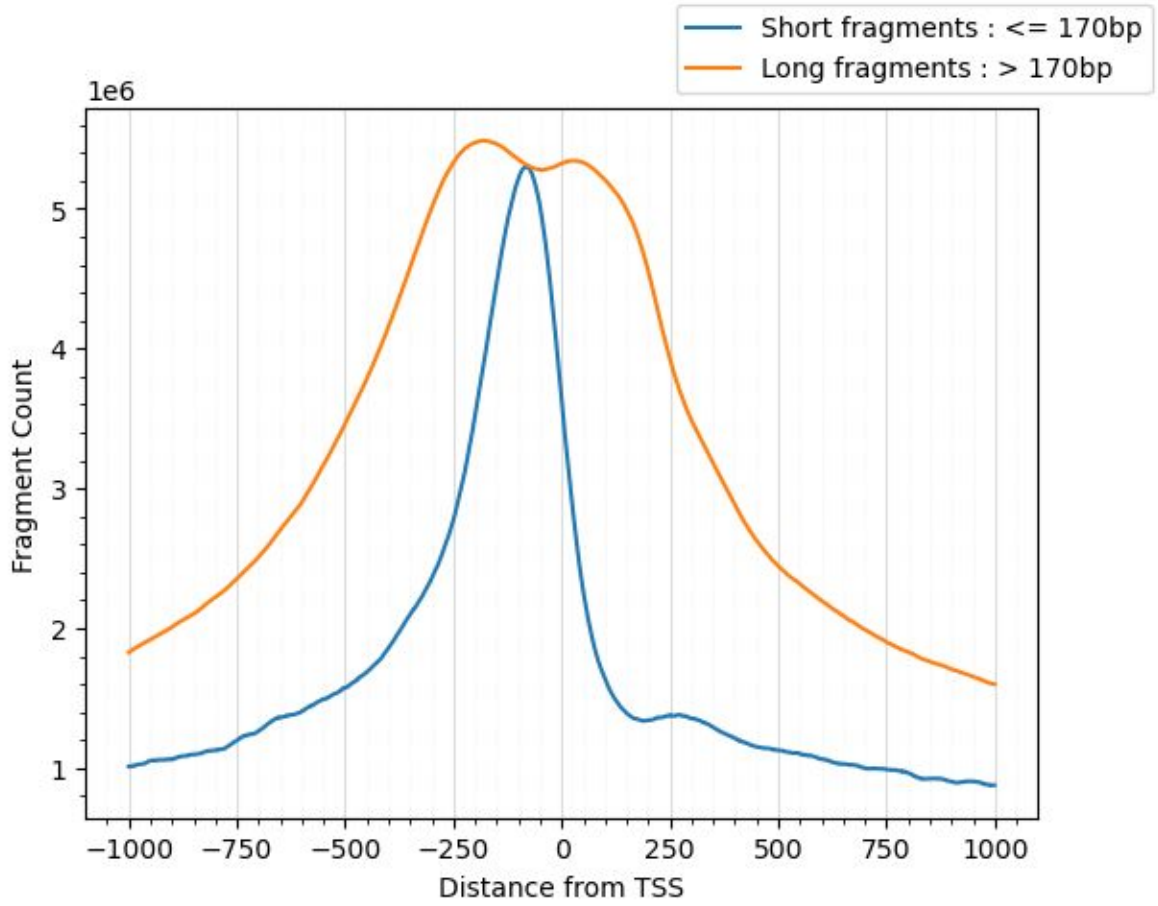
TSS Profile Plots

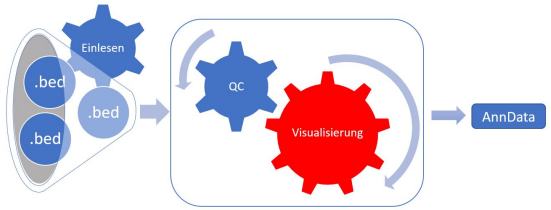
Split Fragmente

Small Intestine



Colon

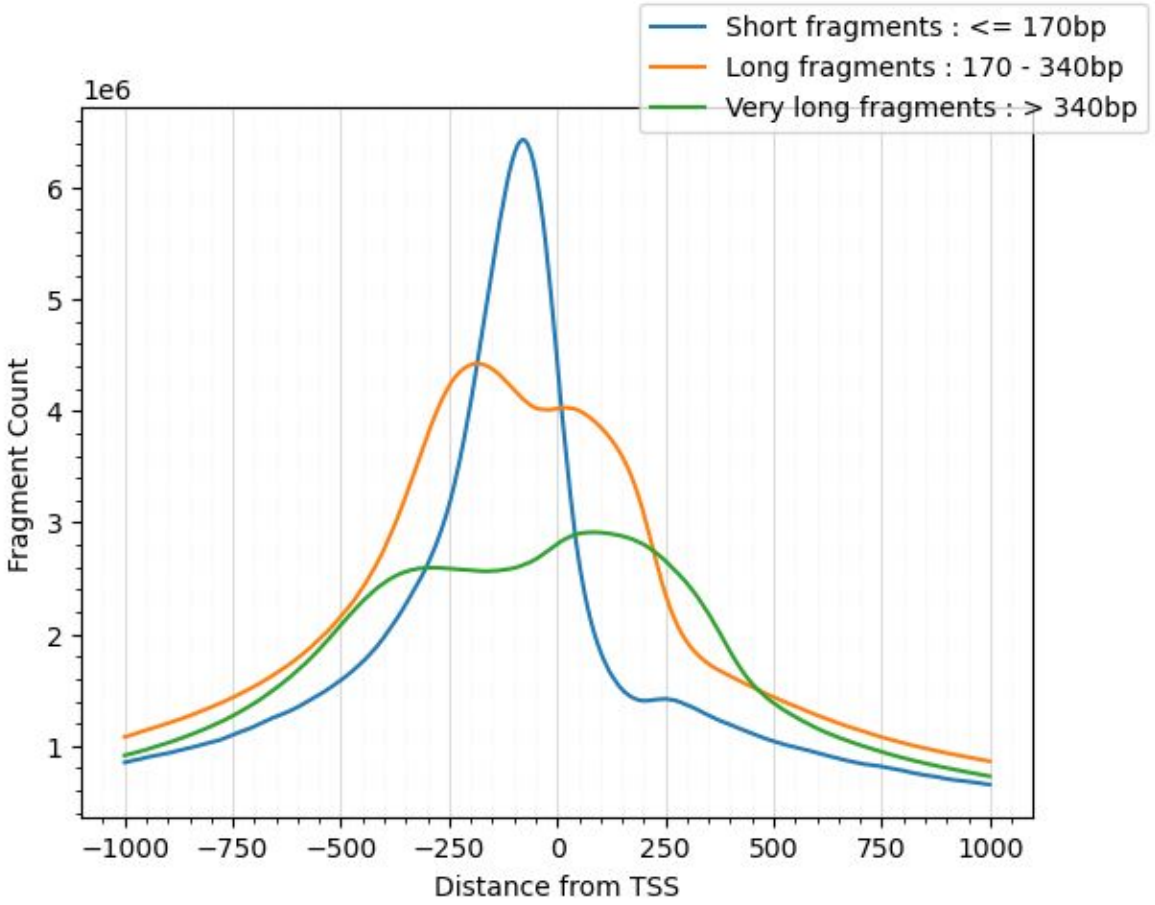




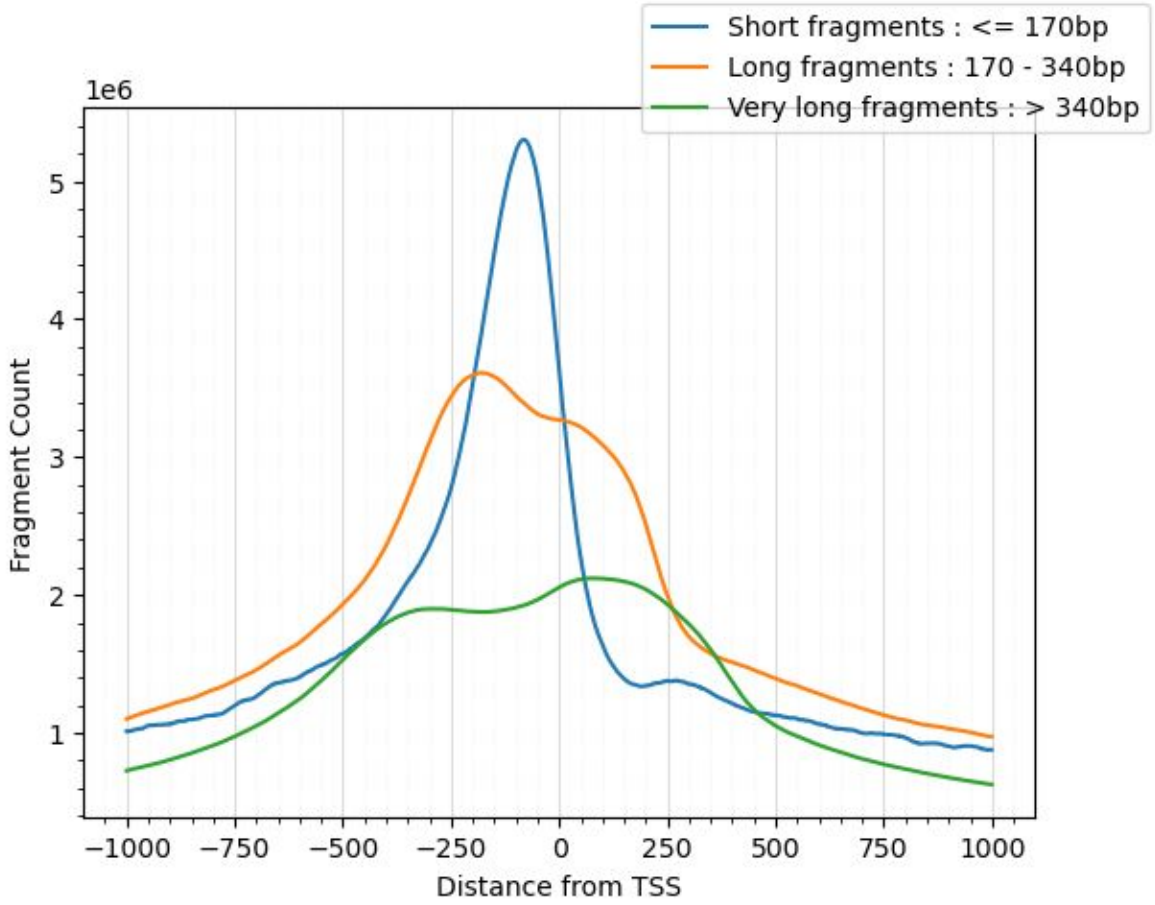
TSS Profile Plots

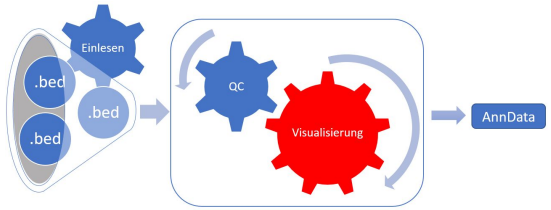
Split Fragmente

Small Intestine



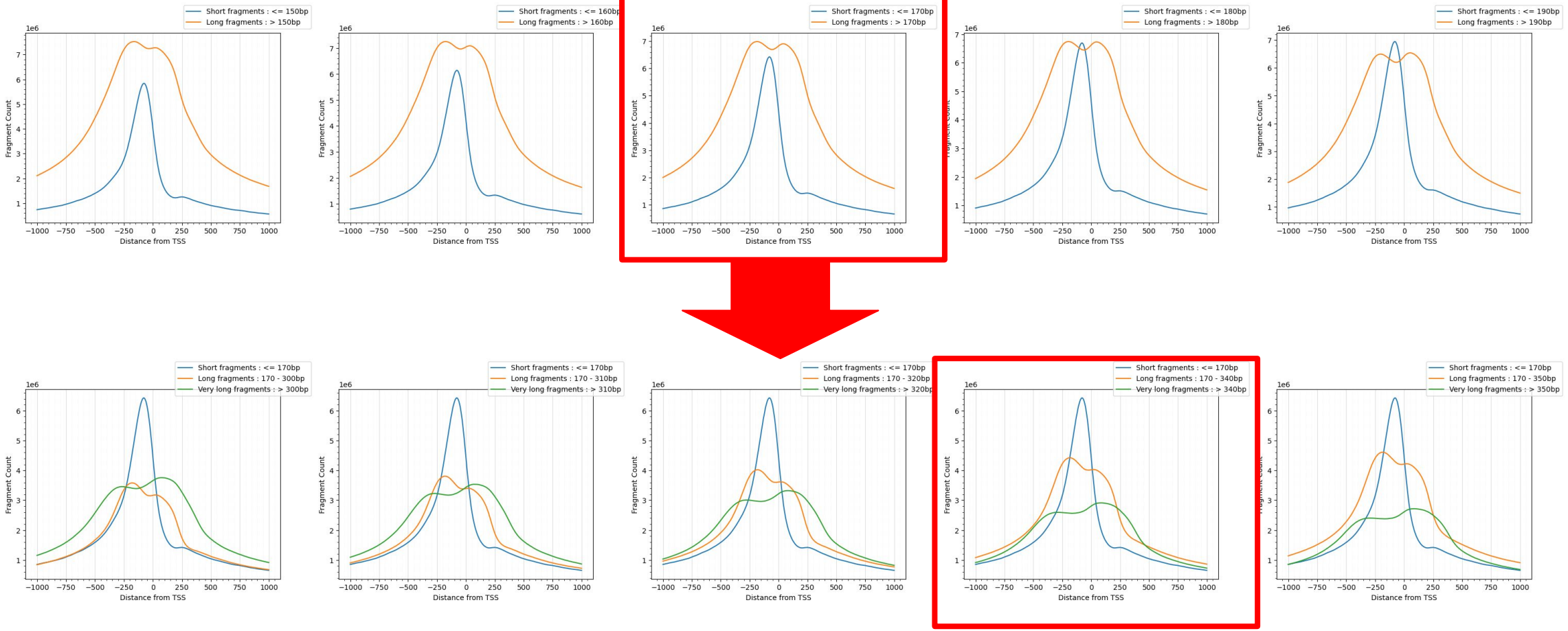
Colon



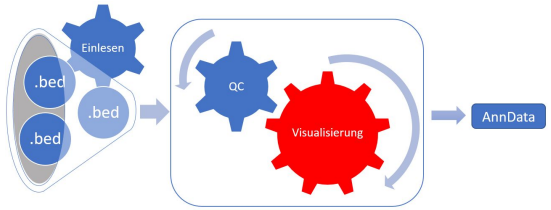


TSS Profile Plots

Multiple Plots



Plot Auswählen



TSS Profile Plots

Vorteile von multiplen Plots:

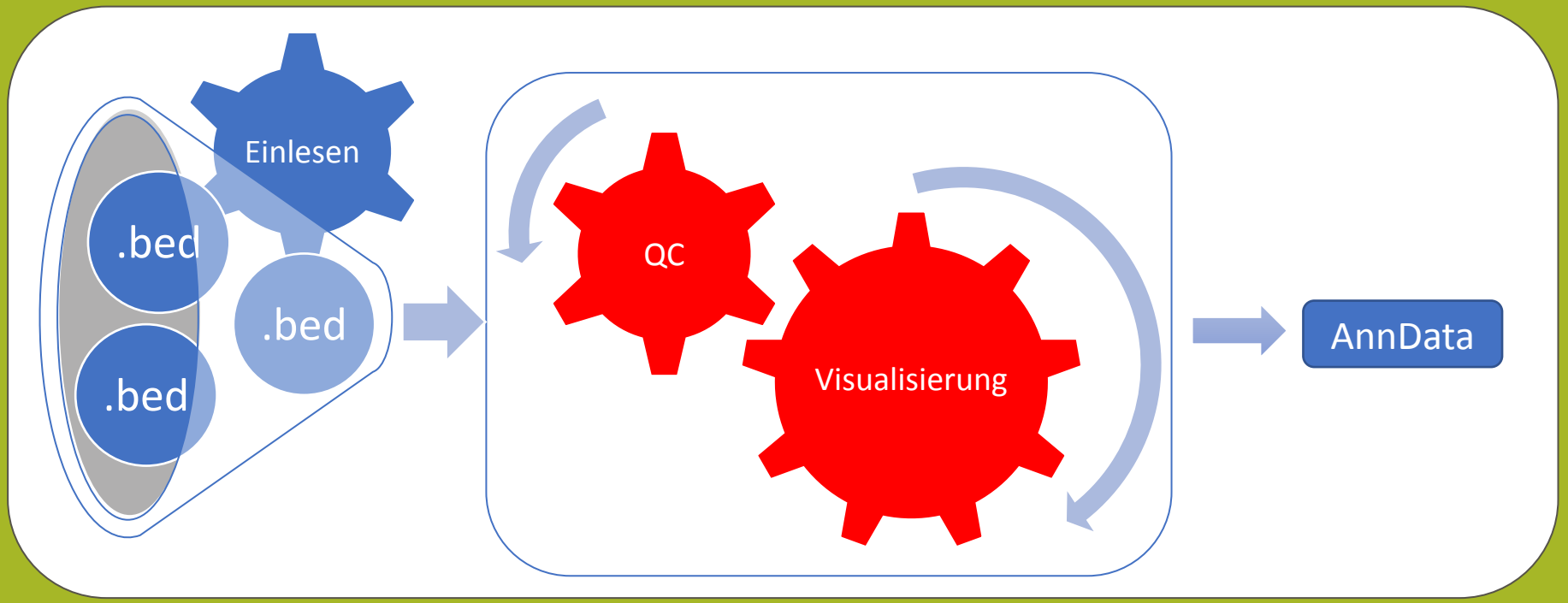
1) Mehr Informationen
2) Einfache Auswahl von Parametern
3) Berechnungsdauer ähnlich wie bei Single Plots

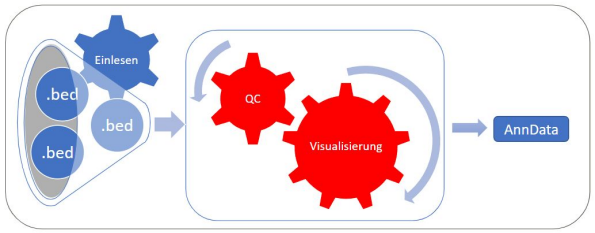
WARTE AUF VERSION 2.0!!!

Features:

- Echtzeit-Visualisierung und Vergleich von Profile Plots im Web

Workflow



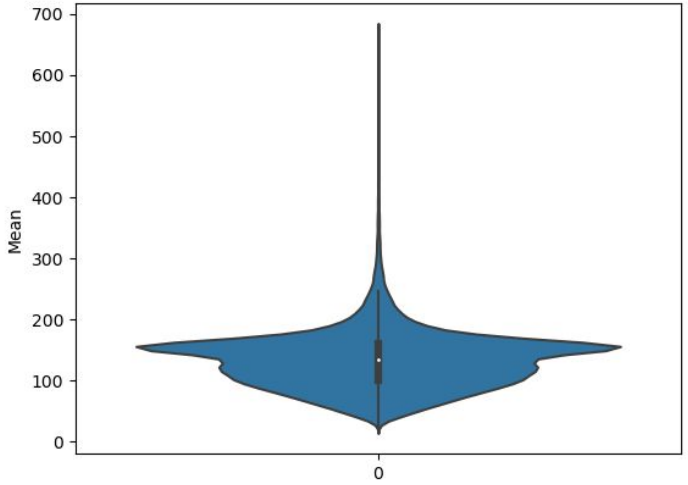


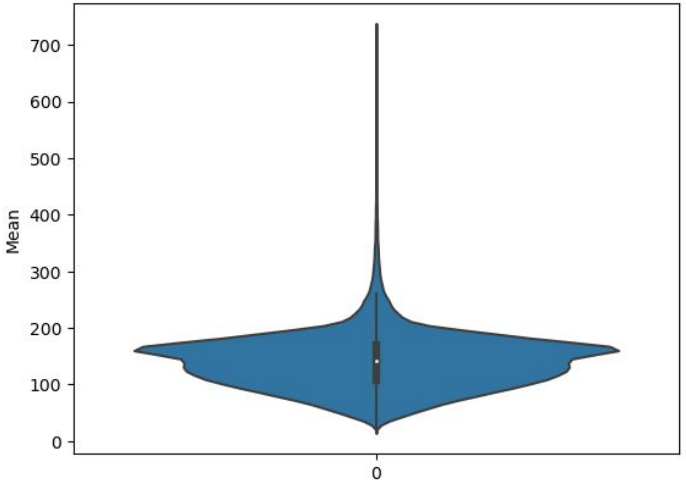
Verteilung der QC-Parameter

Small Intestine

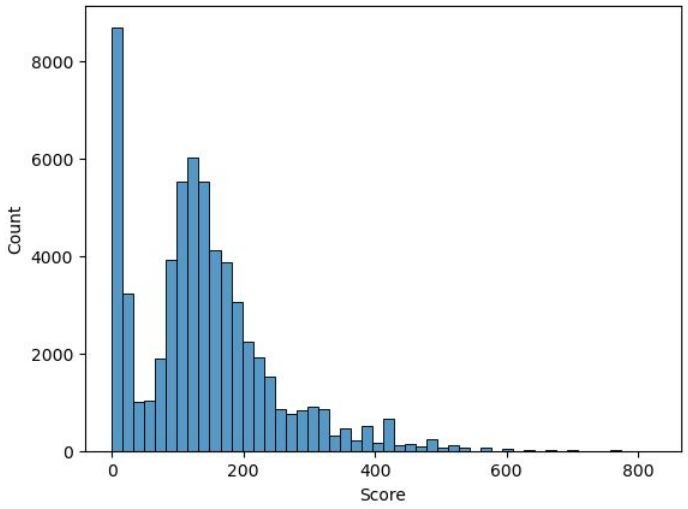
Colon

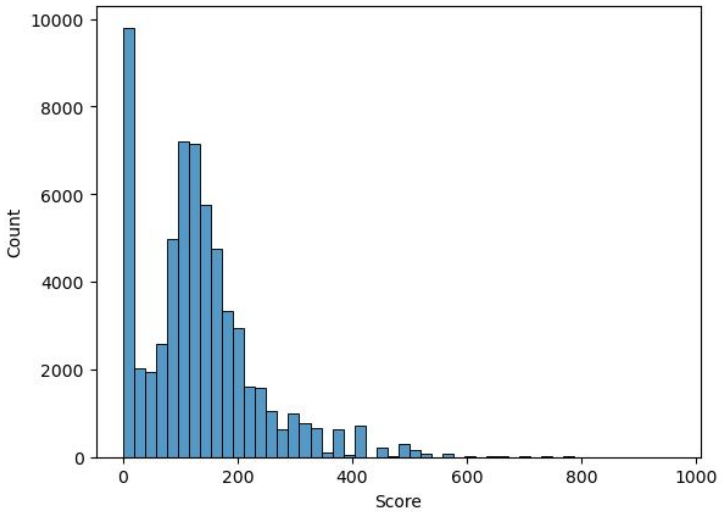
Mean

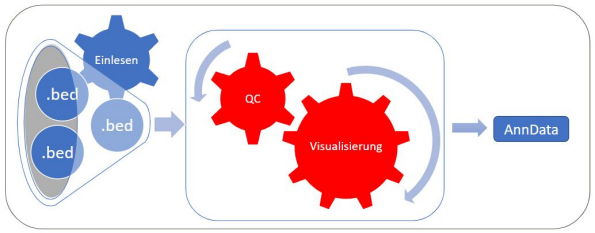




Score





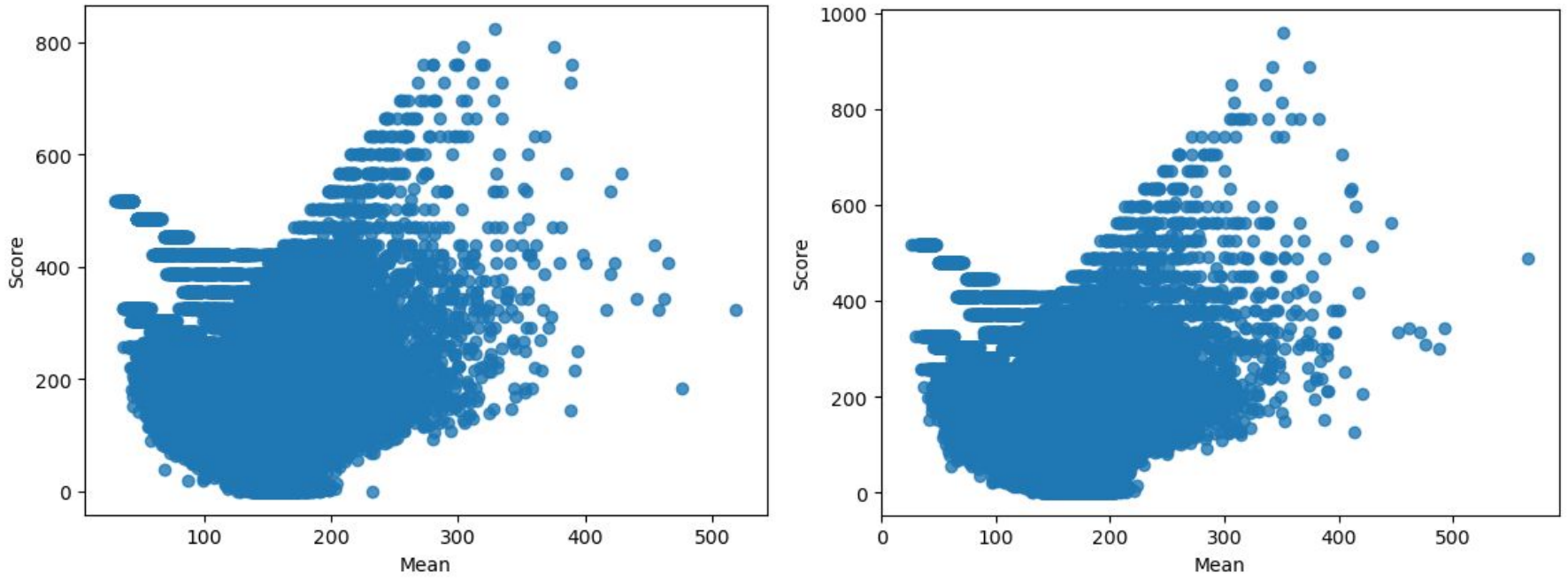


Vergleich QC-Parameter

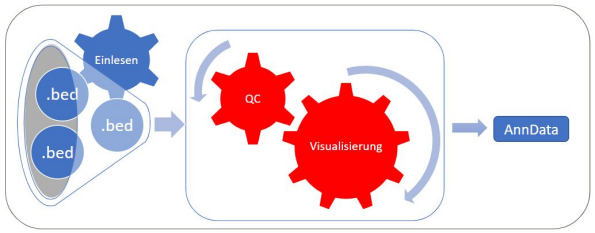
Small Intestine

Colon

Score



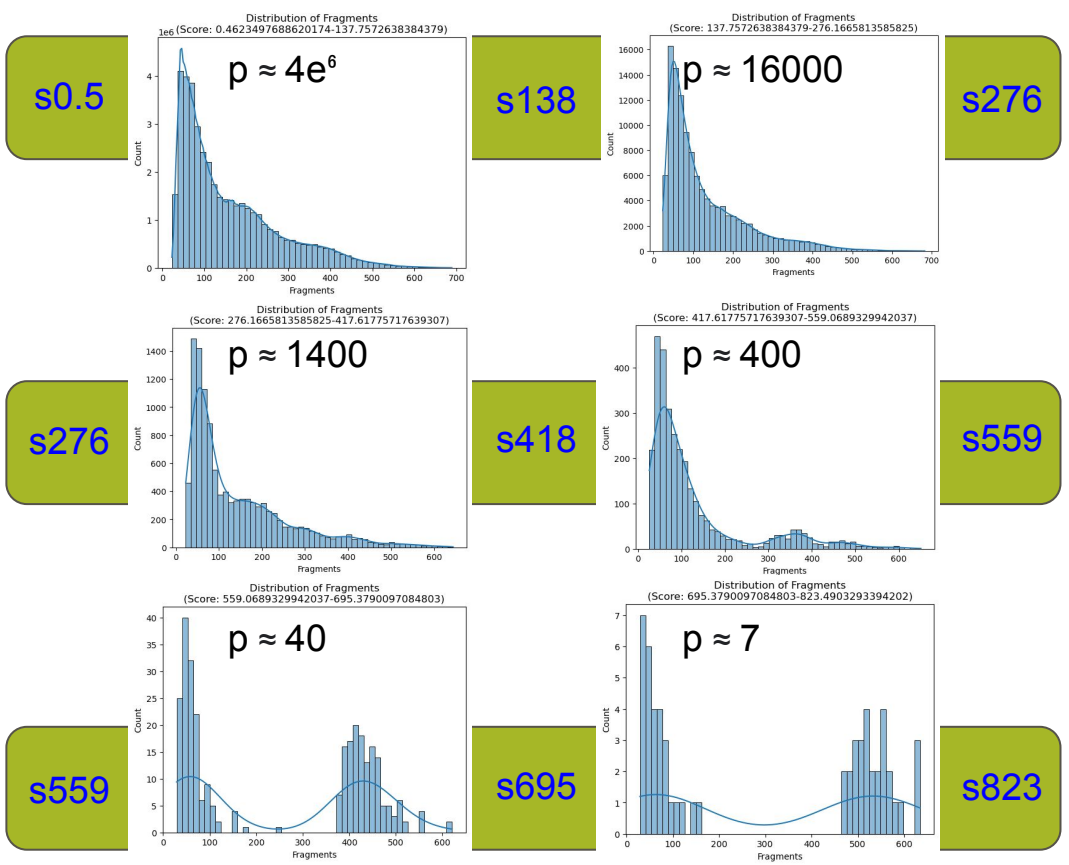
Mean



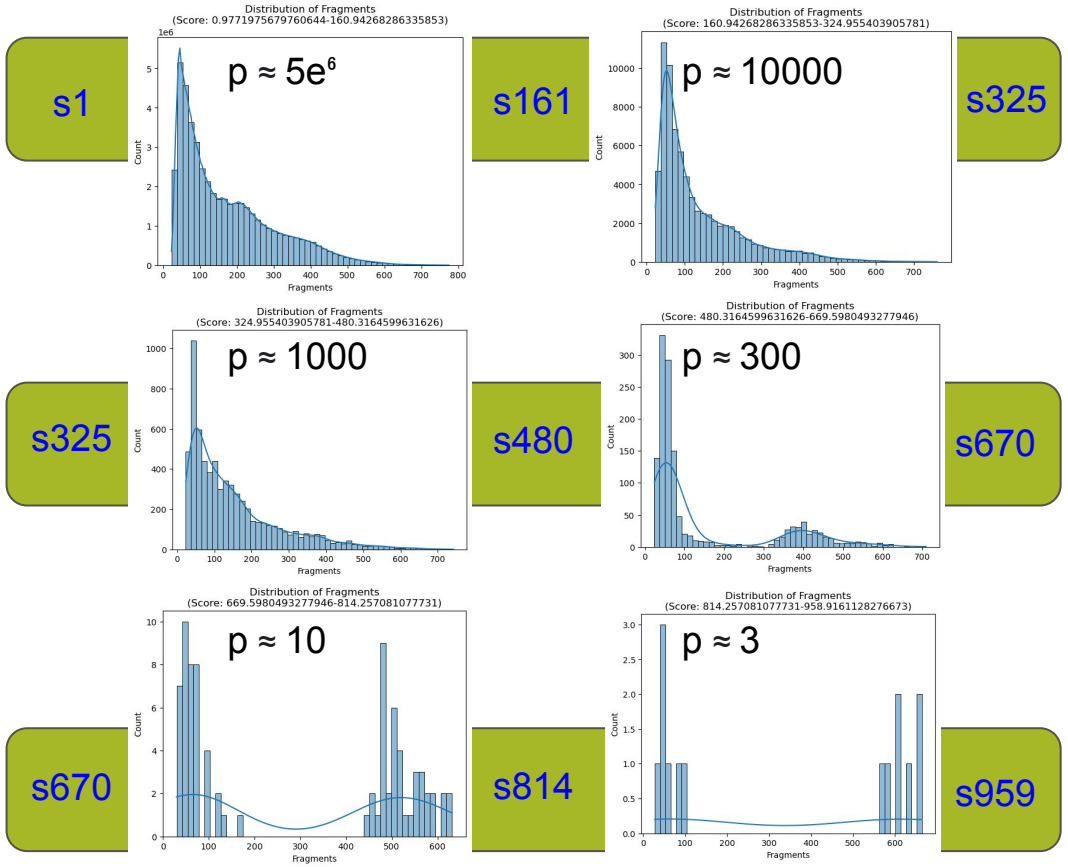
Gruppierung nach Score

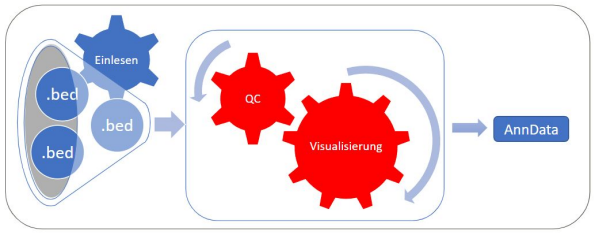
s = Score
p = Peak

Small Intestine



Colon



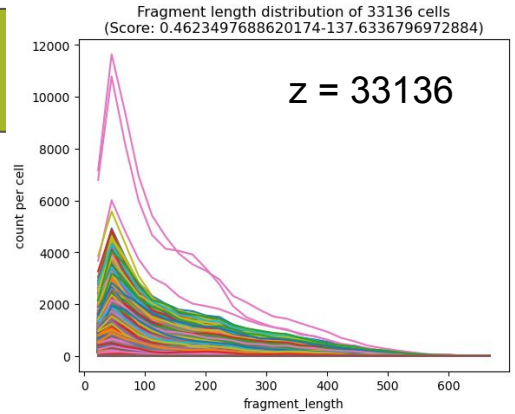


Gruppierung nach Score

s = Score
z = Zellzahl

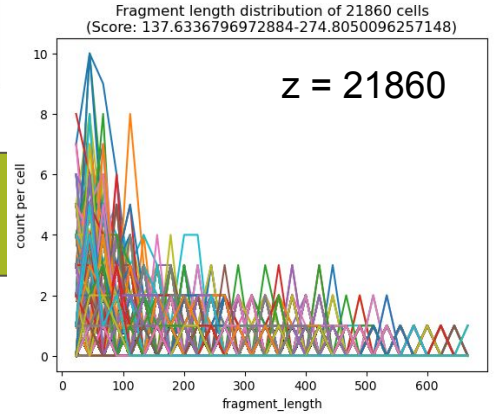
Small Intestine

s0.5



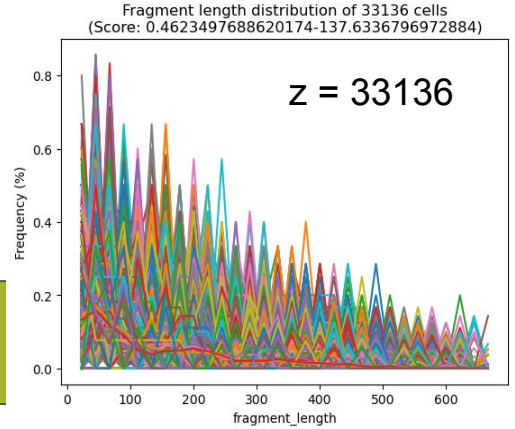
s138

s138



s275

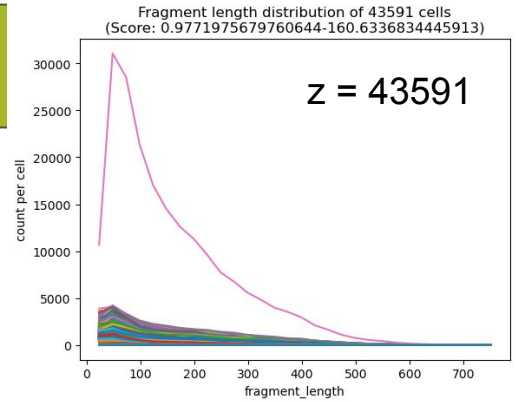
s0.5



s138

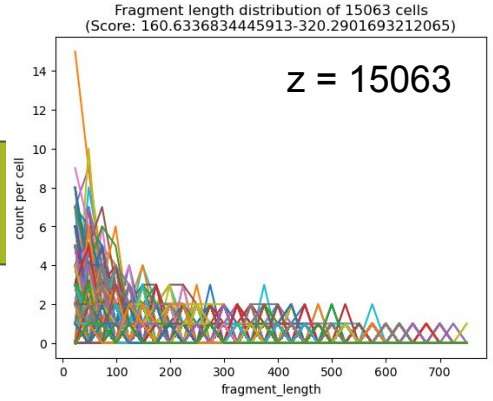
Colon

s0.5



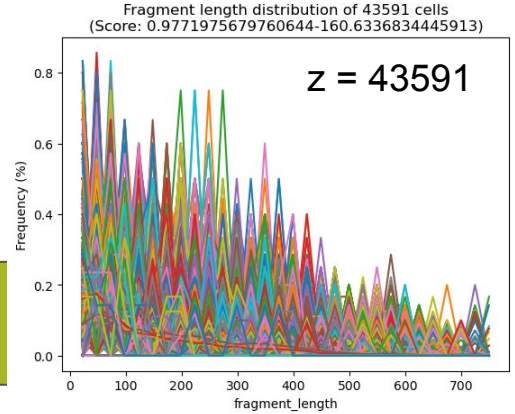
s138

s138

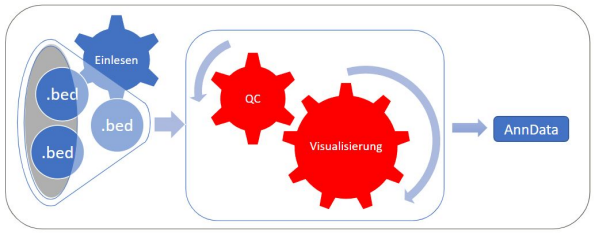


s275

s0.5

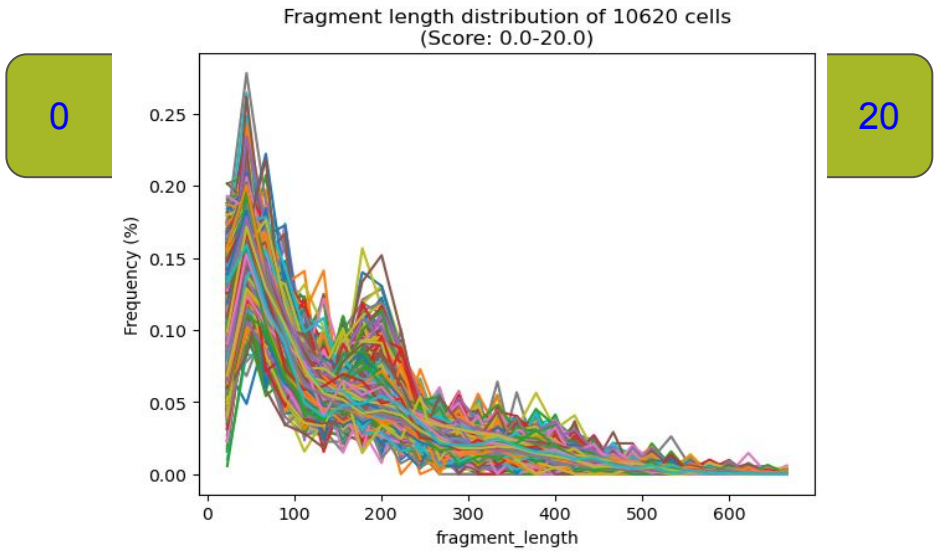


s138

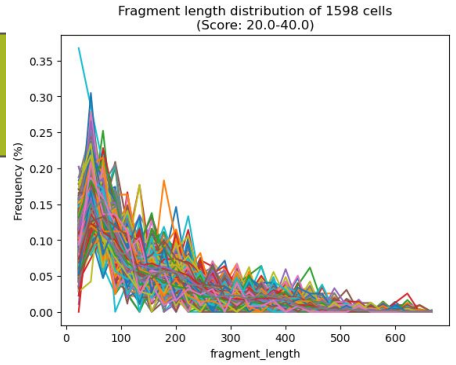


Gruppierung nach Score

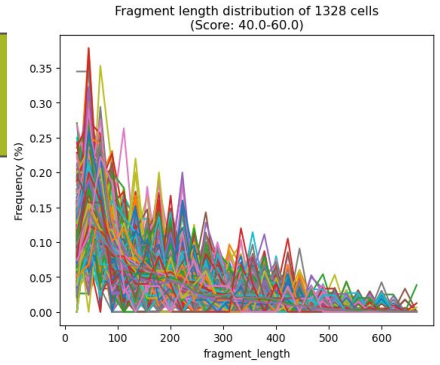
Small Intestine



0 — 20

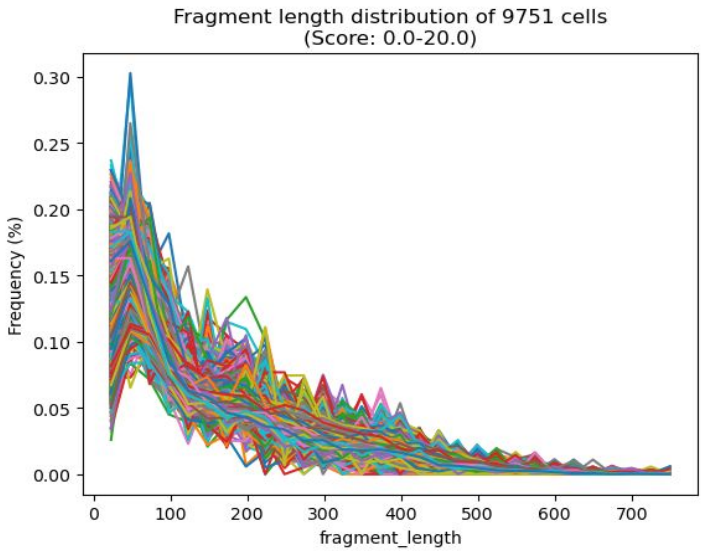


20 — 40

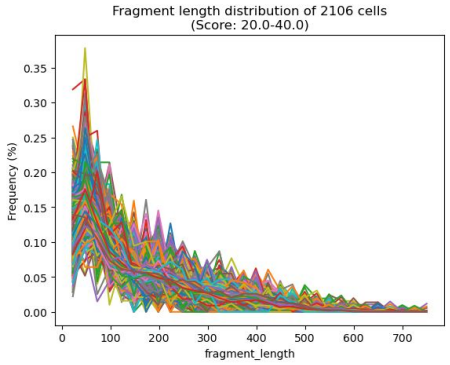


40 — 60

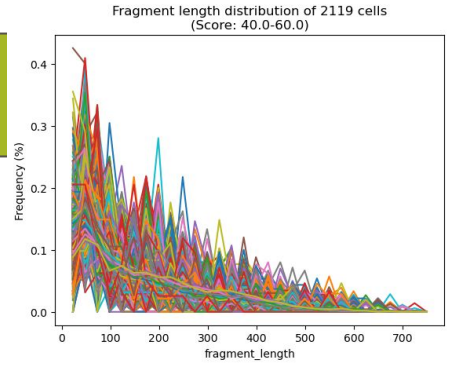
Colon



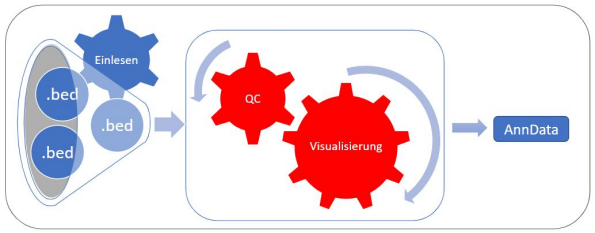
0 — 20



20 — 40



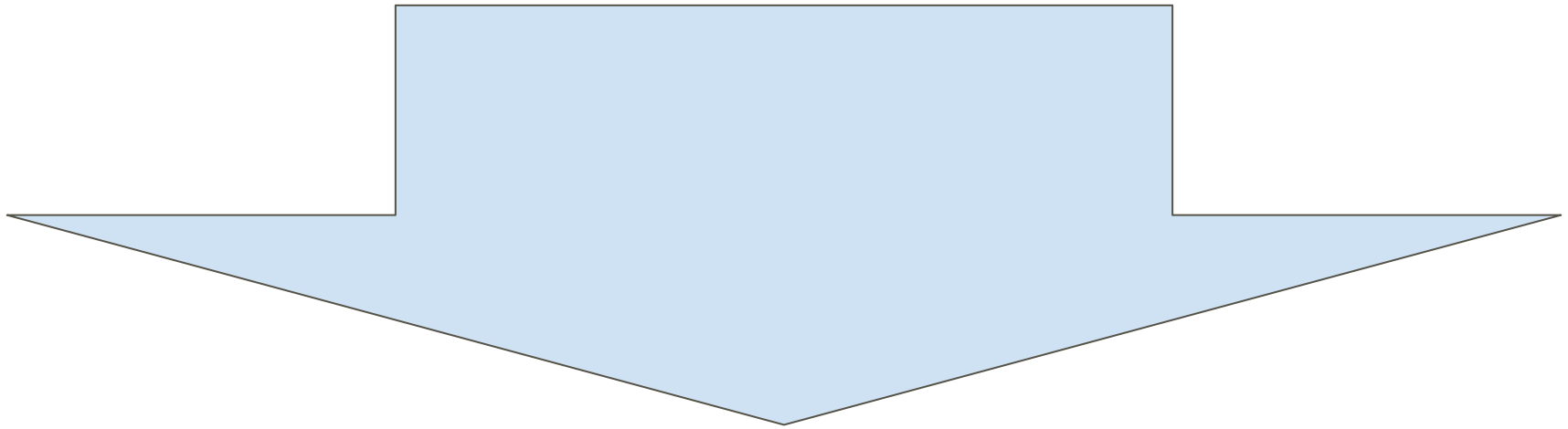
40 — 60



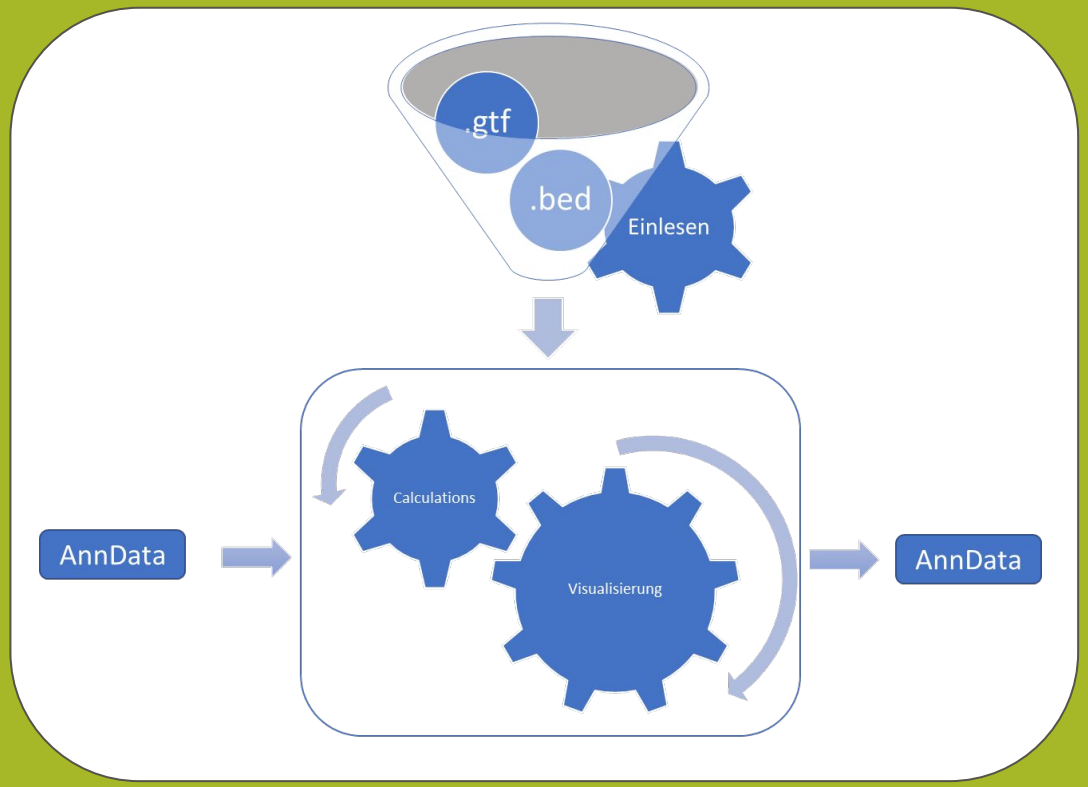
Filtern und Speichern

	Small Intestine	**Colon**
Threshold(s) setzen:	Score = 20	Score = 25
Dataframe filtern:	Alle Zellen mit einem Score < 20	Alle Zellen mit einem Score < 25

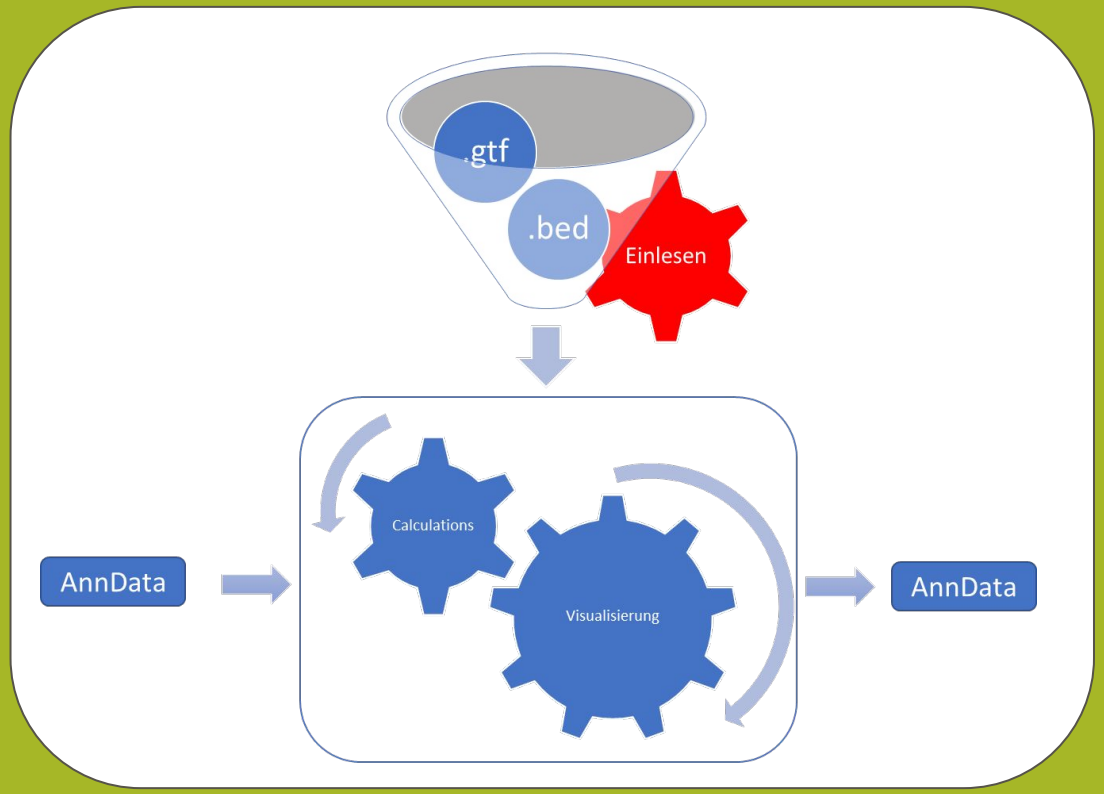
Konvertieren in ein Anndata-Objekt und optional exportieren als .h5ad

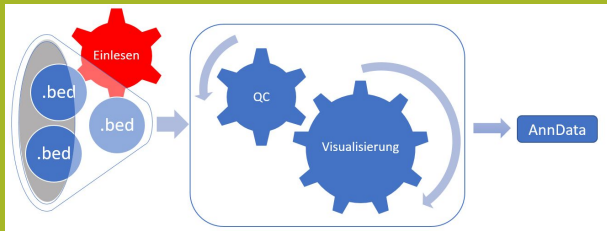


WP2


.gtf
.bed
Einlesen
AnnData
Calculations
Visualisierung
AnnData

Filtern der Referenz Datei



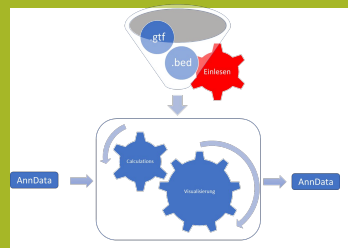


Einlesen der Daten

homo_sapiens.104.mainChr.gtf.bgz

Entpackt

```
chr1    havana  gene    11869   14409   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; gene_name "DDX
chr1    havana  transcript      11869   14409   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transc
chr1    havana  exon    11869   12227   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    12010   12057   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  transcript      12010   13670   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transc
chr1    havana  exon    12179   12227   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    12613   12697   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    12613   12721   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    12975   13052   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    13221   13374   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    13221   14409   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    13453   13670   .       +       .       gene_id "ENSG00000223972"; gene_version "5"; transcript_id
chr1    havana  exon    14404   14501   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  transcript      14404   29570   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transc
chr1    havana  gene    14404   29570   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; gene_name "WAS
chr1    havana  exon    15005   15038   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  exon    15796   15947   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  exon    16607   16765   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  exon    16858   17055   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  exon    17233   17368   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    mirbase transcript      17369   17436   .       -       .       gene_id "ENSG00000278267"; gene_version "1"; transc
chr1    mirbase exon    17369   17436   .       -       .       gene_id "ENSG00000278267"; gene_version "1"; transcript_id
chr1    mirbase gene    17369   17436   .       -       .       gene_id "ENSG00000278267"; gene_version "1"; gene_name "MIR
chr1    havana  exon    17606   17742   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  exon    17915   18061   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
chr1    havana  exon    18268   18366   .       -       .       gene_id "ENSG00000227232"; gene_version "5"; transcript_id
```



Einlesen der Daten

small_intestine_SM-A62GO_rep1_fragments.bed.gz

colon_sigmoid_SM-AZPYO_rep1_fragments.bed.gz

homo_sapiens.104.mainChr.gtf

.bed

.gff

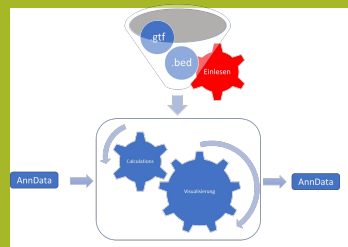
- gene body
- peaks
- enhancer
- blacklisted regions

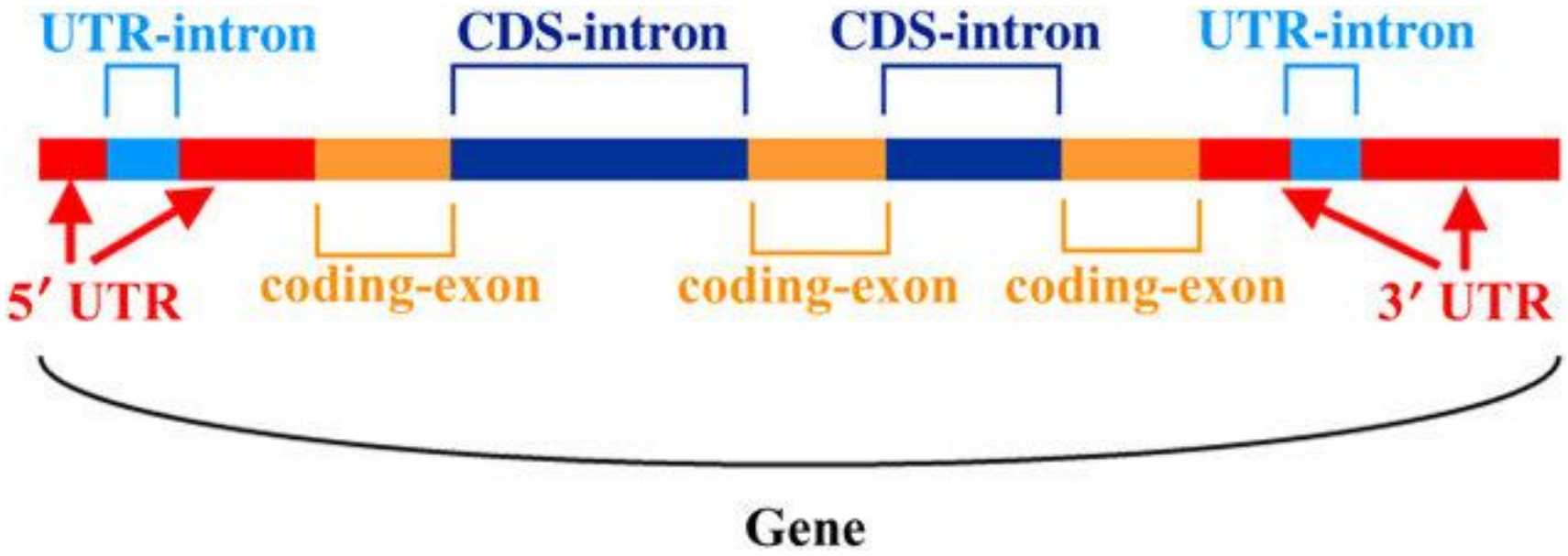
bedtools

gff_analyser

- promoter
- tss
- feature list

feature.gtf





Promotor

- Start Promotor:
  - Start eines Genes minus einer festgelegten Länge
- Ende Promotor:
  - Start des Gens

Enhancer

- Informationen erhältlich in NCBI

TSS

- Start TSS:
  - Start des Genes
- Ende TSS:
  - Start des Genes + TSS Länge

Peaks

- Informationen in anndata.var

Gene Body

- Gene Positionen abzüglich des 5’ und 3’-UTR’s

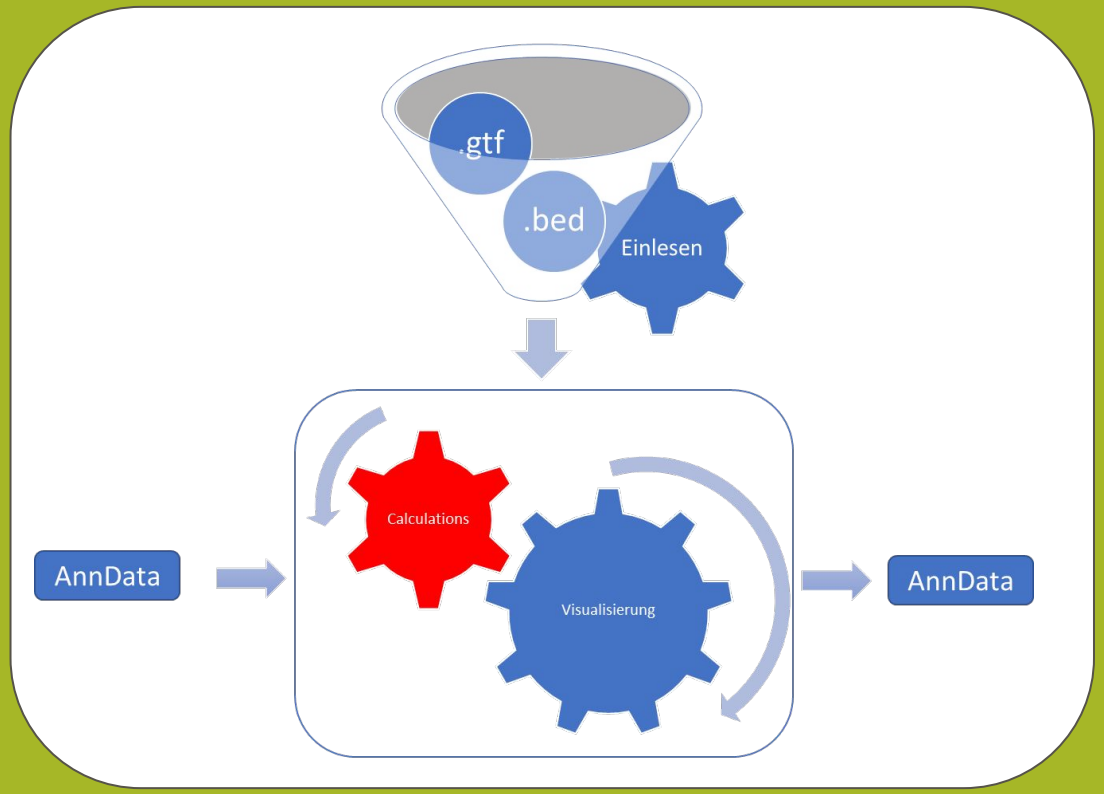
Blacklisted region

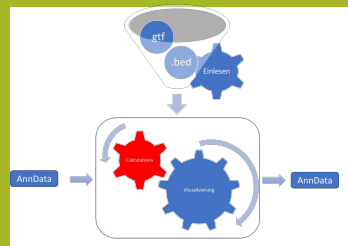
- Informationen erhältlich in NCBI

Feature liste

- In der Referenz Datei annotierte Features der dritten Spalte

Feature Overlap berechnen

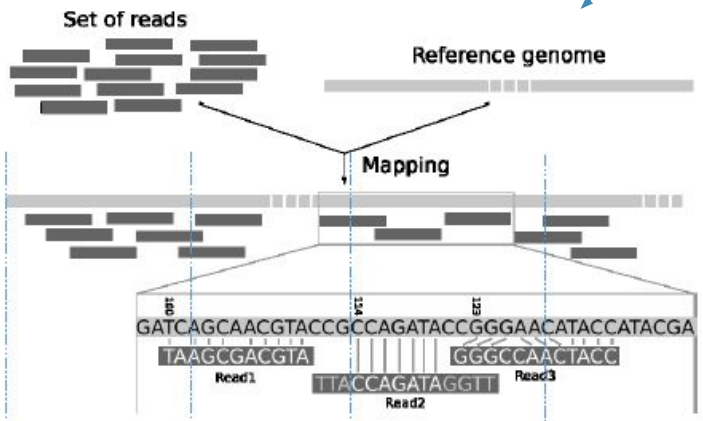




small_intestine_SM-A62GO_rep1_fragments.bed.gz

colon_sigmoid_SM-AZPYO_rep1_fragments.bed.gz

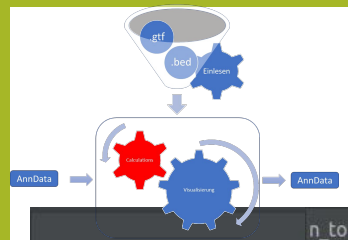
homo_sapiens.104.mainChr.FEATURE.gtf



Feature (peak, promoter etc.)

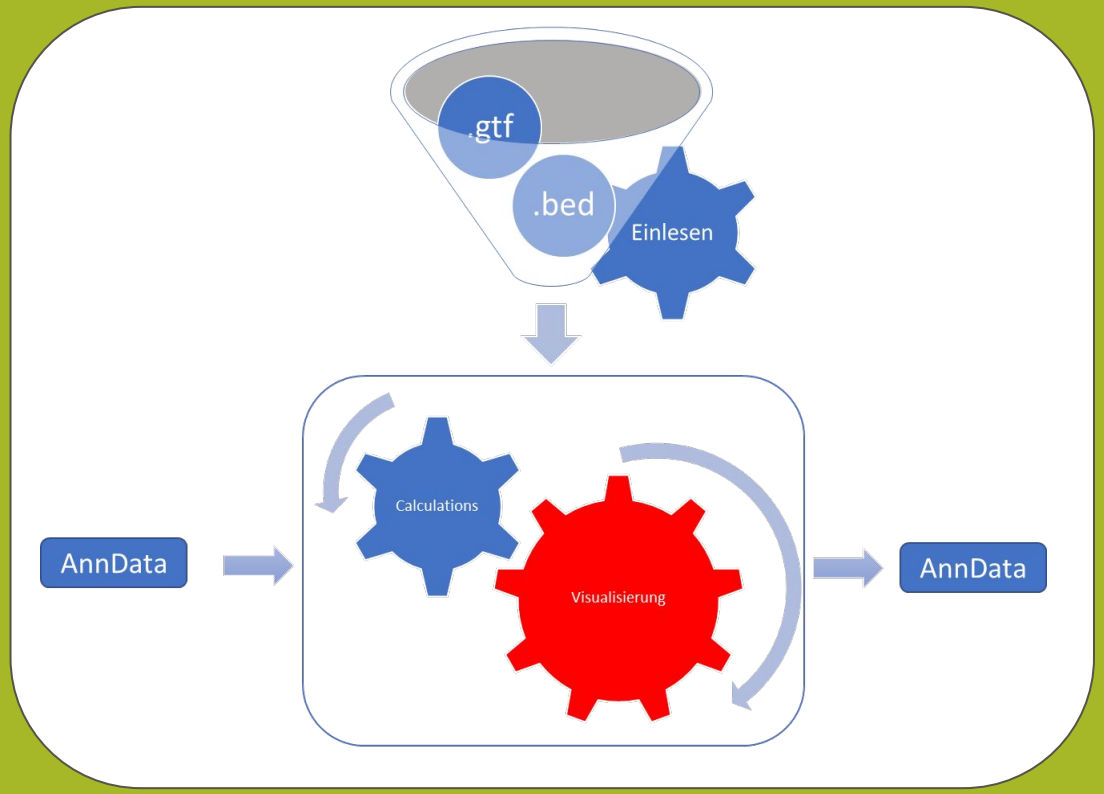
count 6 3

Outside feature count 1 2

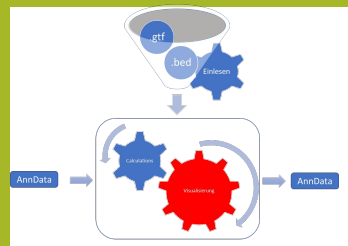


	n_total_fragments	n_fragments_in_TSS100	pct_fragments_in_TSS100	n_fragments_in_gene_bodies	pct_fragments_in_gene_bodies	n_fragments_in_enhancer	pct_fragments_in_enhancer	n_fragments_in_blacklisted	pct_fragments_in_blacklisted
AAACGCAAGCAACCCACTATCT	6640	570	0.08584	4097	0.61702	1651	0.24864	0.00000	0.00000
AAACGCAAGCAAGCGGAACCTA	41644	1586	0.03808	25797	0.61946	10500	0.25214	1.00000	0.00002
AAACGCAAGCAAGCTCTCTGAT	19923	894	0.04487	12002	0.60242	4146	0.20810	6.00000	0.00030
AAACGCAAGCAAGGGCAAATCA	10955	762	0.06956	7177	0.65513	2551	0.23286	3.00000	0.00027
AAACGCAAGCAATATAGGTCCG	9921	1068	0.10765	6808	0.68622	2286	0.23042	3.00000	0.00030
AAACGCAAGCACGGGTTCTTGT	10389	862	0.08297	6625	0.63769	2591	0.24940	0.00000	0.00000
AAACGCAAGCAGGAACGCCGAT	17421	1662	0.09540	11953	0.68613	2934	0.16842	9.00000	0.00052
AAACGCAAGCAGGGATGAGTAA	3978	209	0.05254	2429	0.61061	923	0.23203	0.00000	0.00000
AAACGCAAGCAGGTCACGACTA	7679	495	0.06446	4994	0.65035	2058	0.26800	0.00000	0.00000
AAACGCAAGCATAAGGTCCTGA	21907	1662	0.07587	14744	0.67303	4537	0.20710	8.00000	0.00037
AAACGCAAGCATCGAGGCCTAA	10996	1271	0.11559	7761	0.70580	1915	0.17415	2.00000	0.00018
AAACGCAAGCATTTGGCCTCTA	58962	10654	0.18069	47648	0.80811	5765	0.09777	122.00000	0.00207
AAACGCAAGCCAGCGAAGACCT	9176	830	0.09045	6141	0.66925	1488	0.16216	6.00000	0.00065
AAACGCAAGCCCTCATAGTAGA	7502	805	0.10730	4655	0.62050	1809	0.24114	0.00000	0.00000
AAACGCAAGCCGATTCCGGCAT	11522	641	0.05563	7724	0.67037	2510	0.21784	1.00000	0.00009
AAACGCAAGCCGTCCCTAGAAC	7787	622	0.07988	4600	0.59073	1704	0.21883	3.00000	0.00039
AAACGCAAGCCTCCCGTGTAAG	21846	3037	0.13902	15800	0.72324	4051	0.18543	13.00000	0.00060
AAACGCAAGCGAAGGACCTAGT	21279	3406	0.16006	15974	0.75069	2900	0.13628	26.00000	0.00122
AAACGCAAGCGAGATAGTACTG	19901	1830	0.09196	13092	0.65786	3848	0.19336	14.00000	0.00070
AAACGCAAGCGATACGCCGTAG	2506	191	0.07622	1616	0.64485	665	0.26536	0.00000	0.00000
AAACGCAAGCGCAGCTGATGTG	15942	1237	0.07759	10179	0.63850	4233	0.26553	0.00000	0.00000
AAACGCAAGCGCCTAAACGAAG	4073	408	0.10017	2506	0.61527	995	0.24429	0.00000	0.00000
AAACGCAAGCGCTAGTCAAGAG	11024	1177	0.10677	7118	0.64568	2541	0.23050	6.00000	0.00054
AAACGCAAGCGGACTTCTAGTT	9737	609	0.06254	6015	0.61775	2598	0.26682	0.00000	0.00000
AAACGCAAGCGGATAAAGGGAT	12758	1518	0.11898	8427	0.66053	2087	0.16358	17.00000	0.00133
AAACGCAAGCGGCACTTCATCT	10730	1024	0.09543	6604	0.61547	2552	0.23784	0.00000	0.00000
AAACGCAAGCGGGCAGTGGTAA	20350	2344	0.11518	13767	0.67651	4191	0.20595	12.00000	0.00059
AAACGCAAGCGGTGTCGGTTGT	14167	1392	0.09826	9243	0.65243	3244	0.22898	5.00000	0.00035
AAACGCAAGCGTACAACTCTAG	9855	488	0.04952	6272	0.63643	1815	0.18417	6.00000	0.00061
AAACGCAAGCGTTGGTCTACTA	6946	621	0.08940	4321	0.62208	1765	0.25410	0.00000	0.00000
AAACGCAAGCTAACGCGTACCC	22966	2548	0.11095	14287	0.62209	5702	0.24828	2.00000	0.00009
AAACGCAAGCTACCTCCAACTT	4086	189	0.04626	2446	0.59863	781	0.19114	0.00000	0.00000
AAACGCAAGCTAGATCCACGTA	4660	419	0.08991	3046	0.65365	763	0.16373	3.00000	0.00064
AAACGCAAGCTCCGTAAGGTTC	14182	1584	0.11169	8666	0.61106	3535	0.24926	1.00000	0.00007
AAACGCAAGCTGGAGTAACCAT	8980	875	0.09744	5613	0.62506	2255	0.25111	2.00000	0.00022

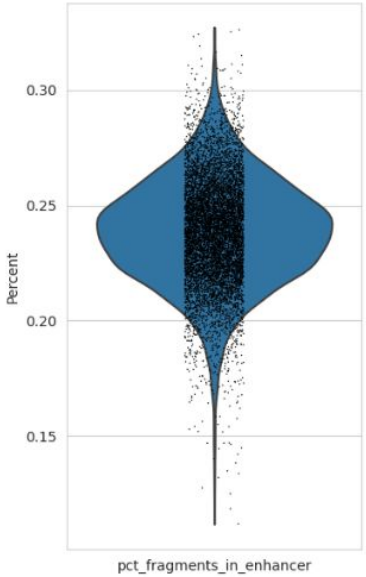
Visualisierung

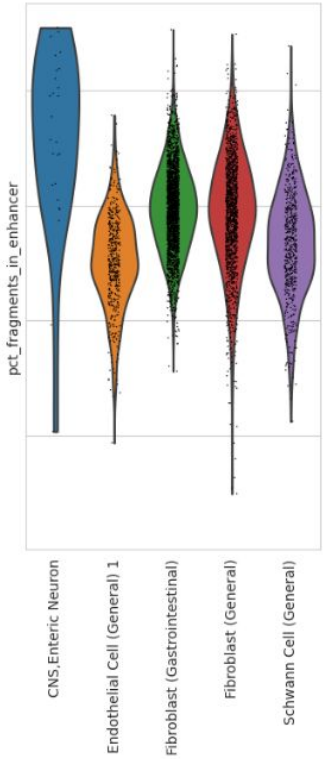


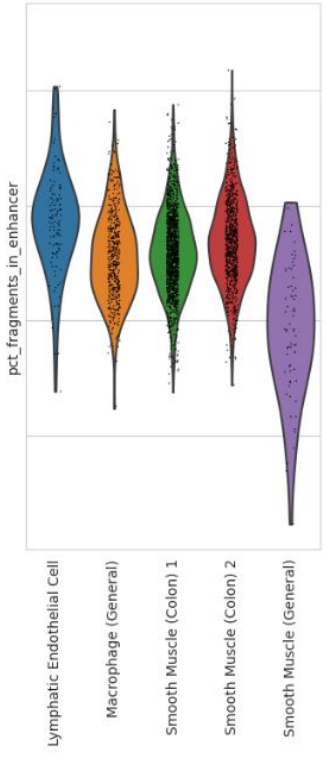
Scatter/Violin-Plots

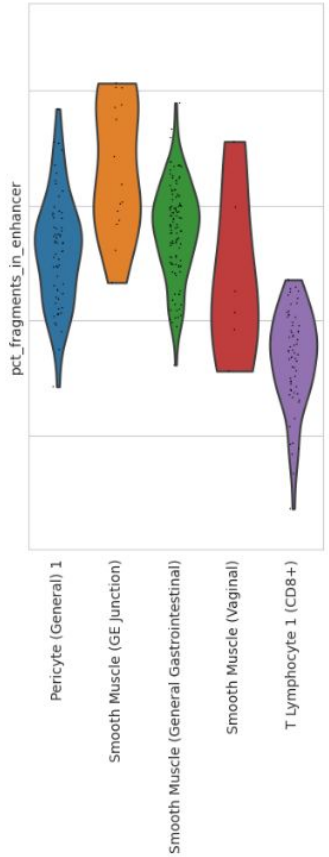


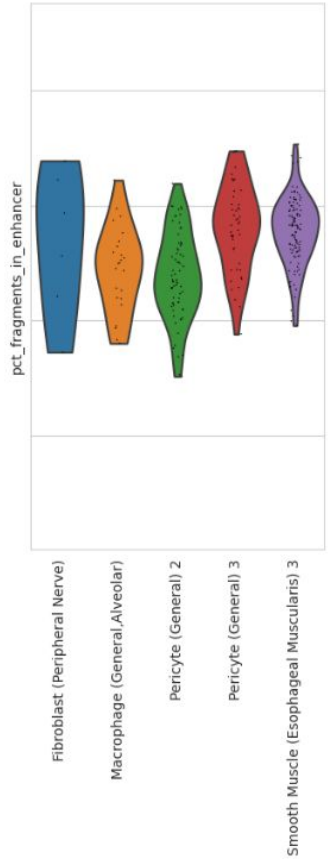
Violinen Plots - Colon

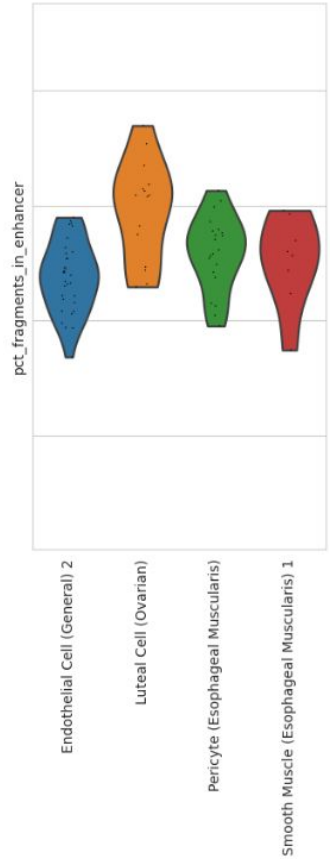


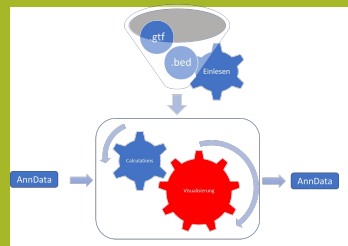




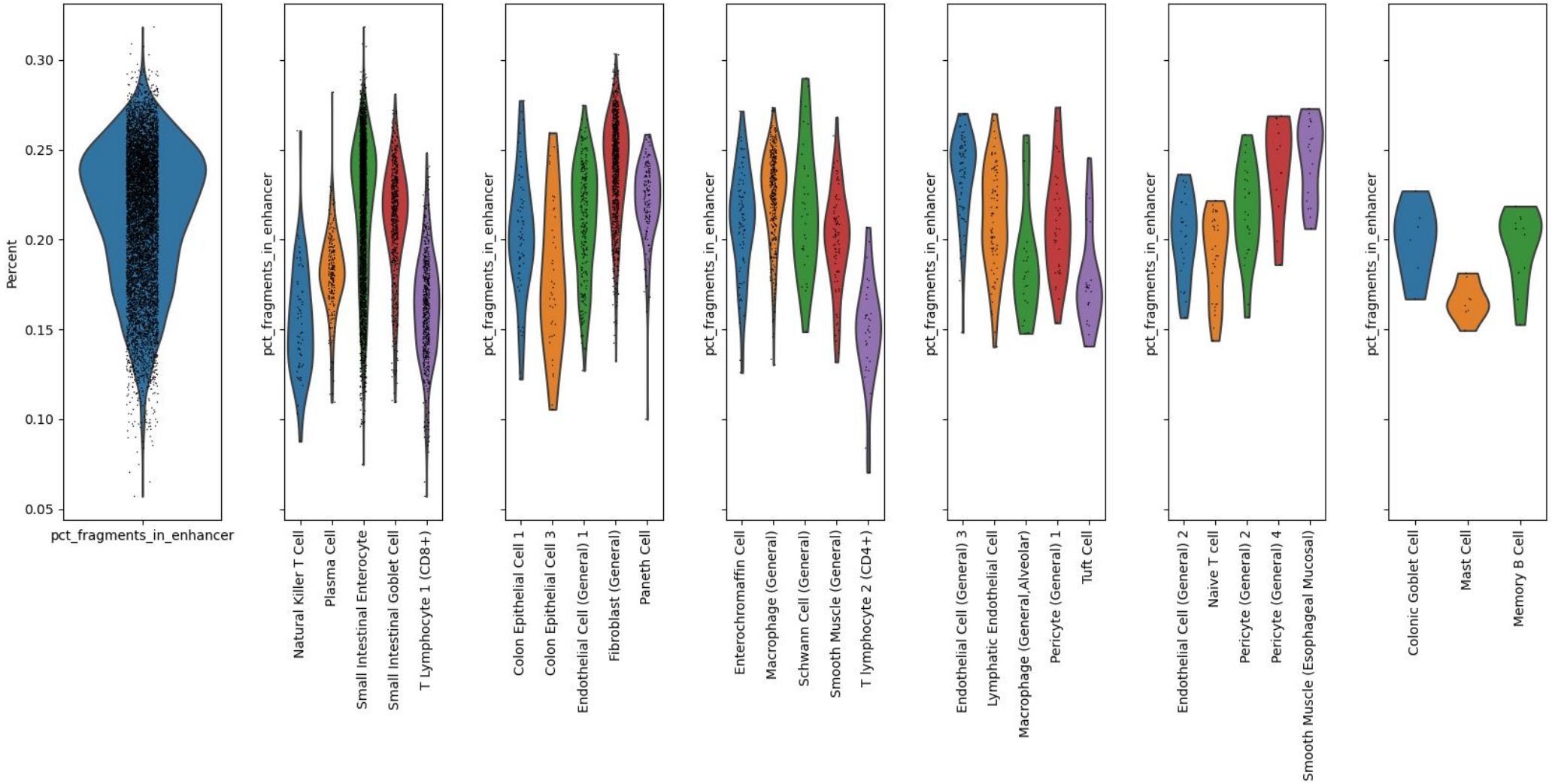


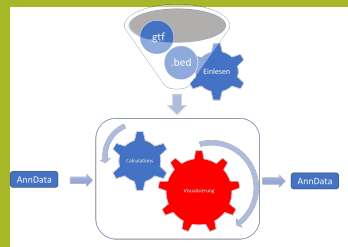






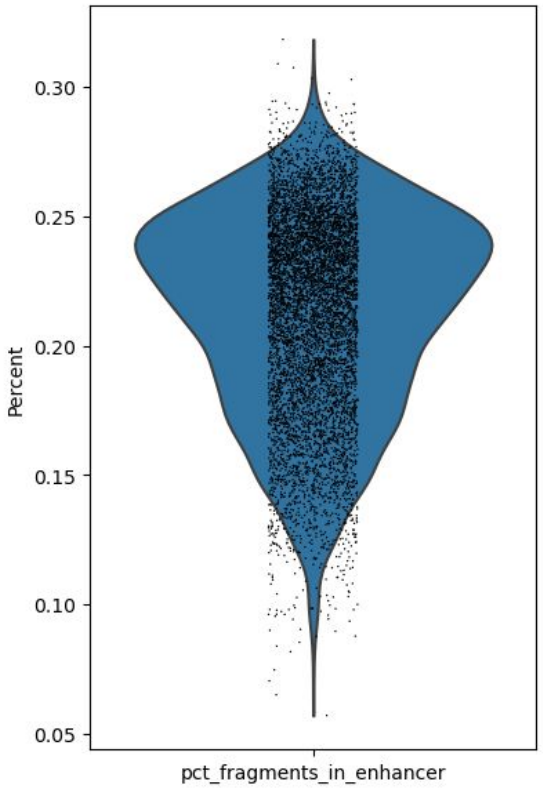
Violinen Plots - Intestinal



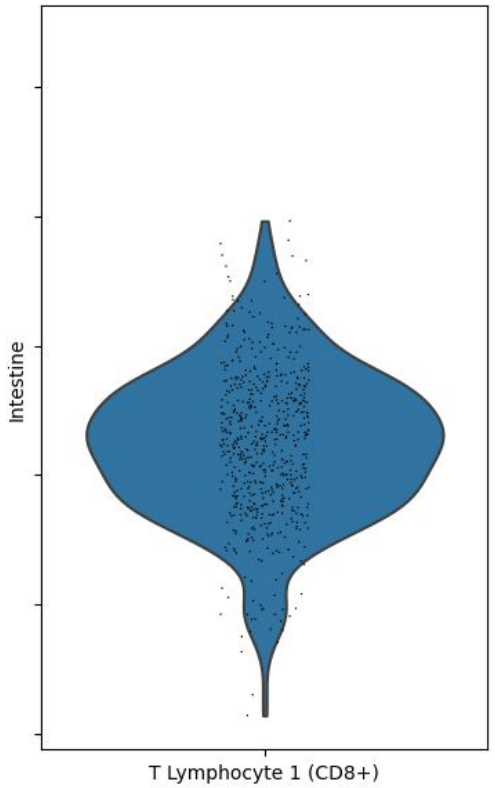


Violinen Plots CD8+ Zellen - Enhancer

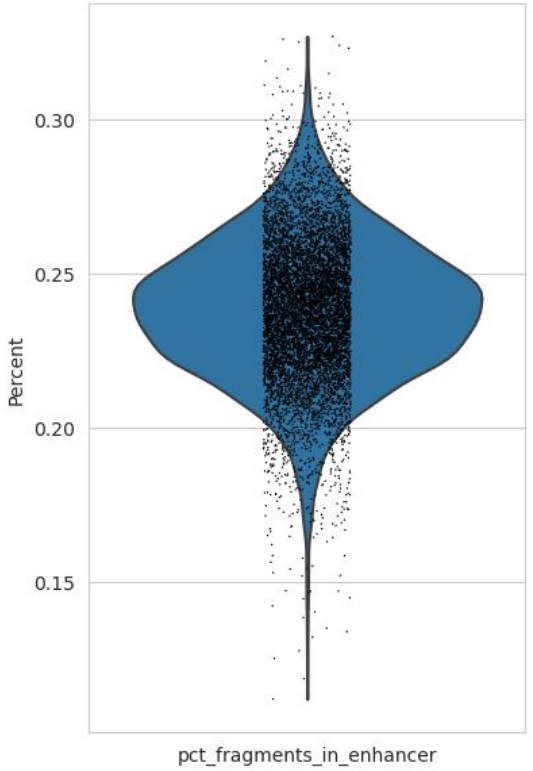
Zellen insgesamt: 7470
Lymphozyte (CD8+): 616



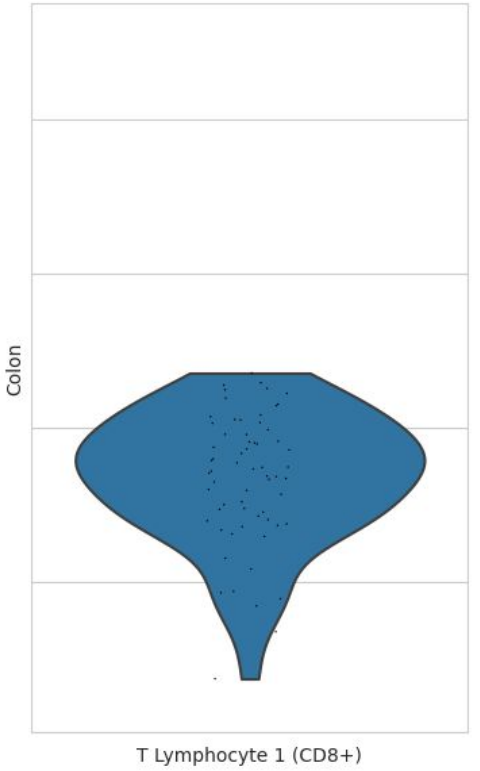
Intestine

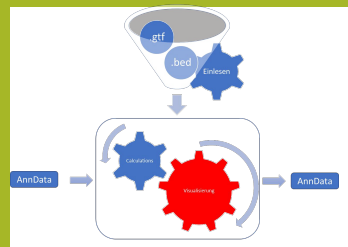


Colon

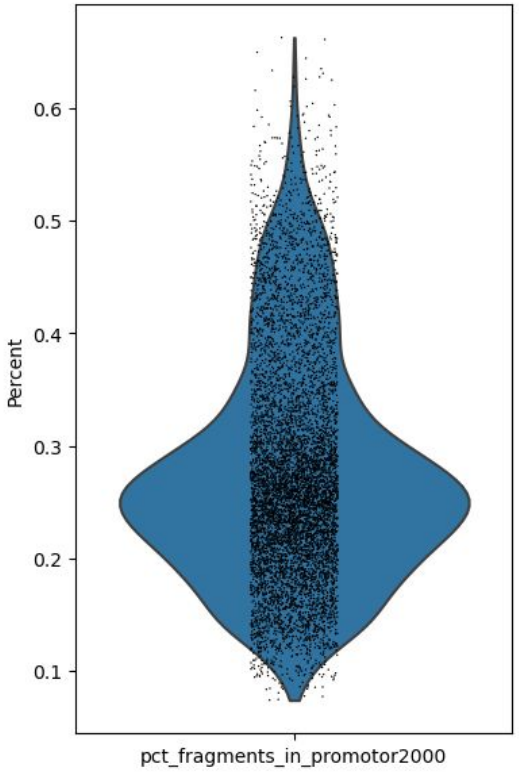


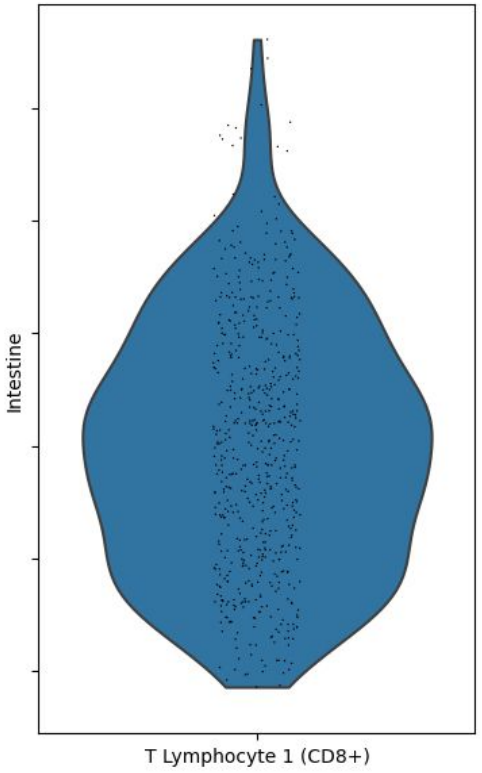
Zellen instgesamt: 8682
Lymphozyten (CD8+): 64

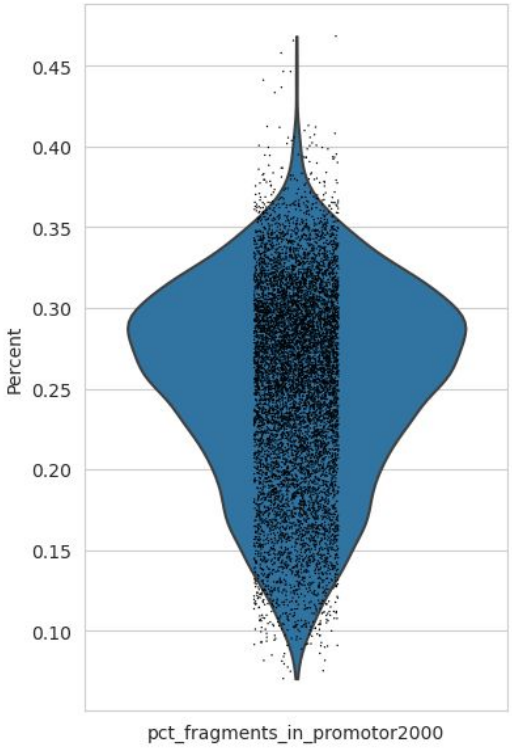


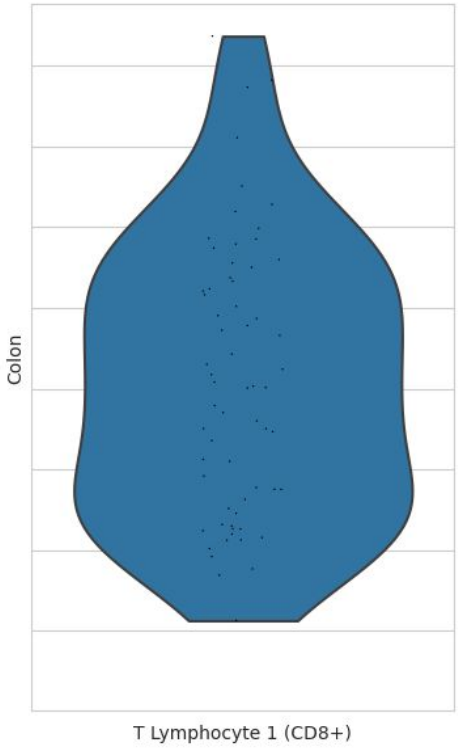


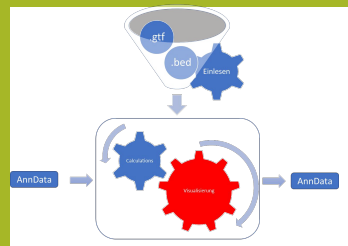
Violinen Plots CD8+ Zellen - Promotor



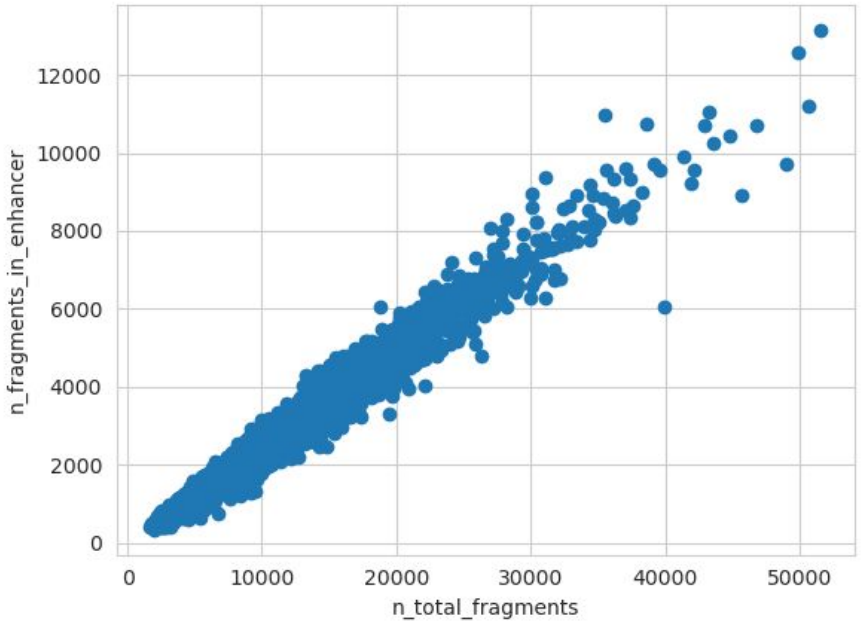


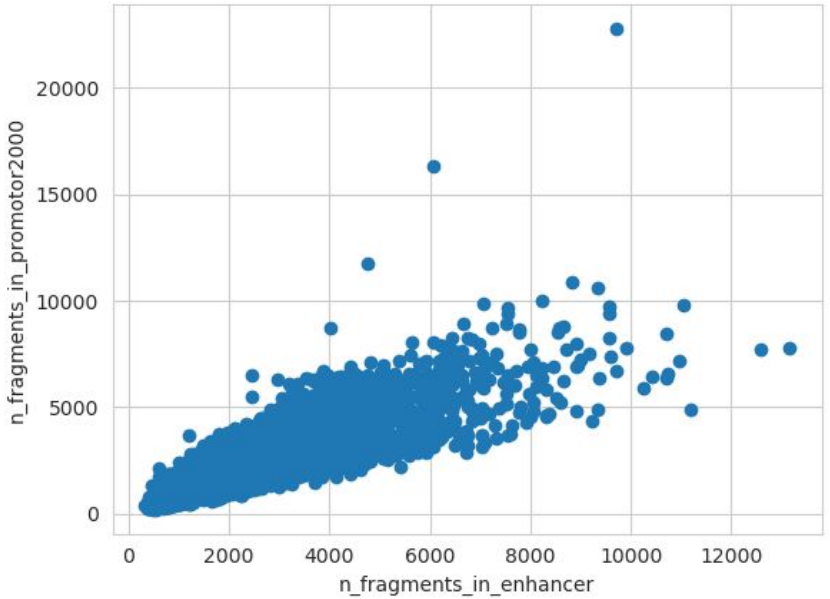


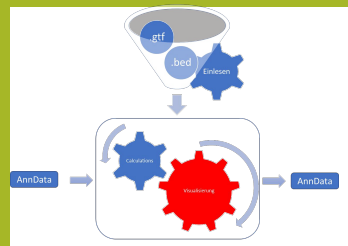




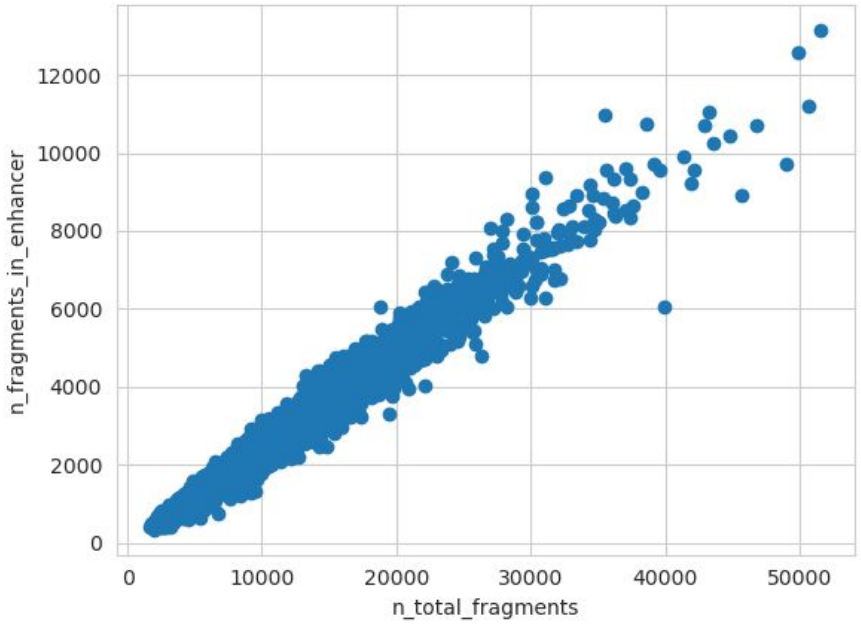
Scatter Plots - Colon

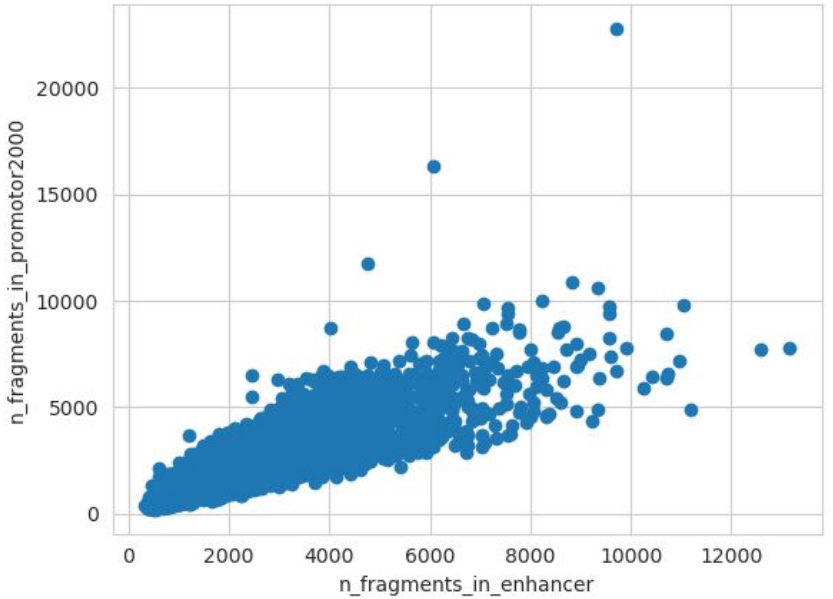




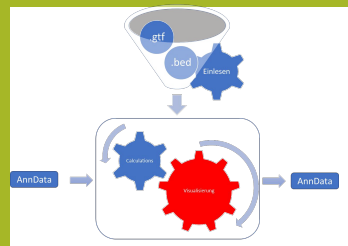


Scatter Plots - Intestine

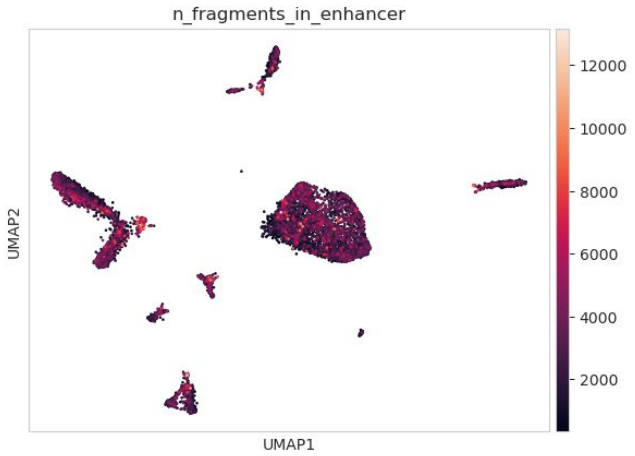


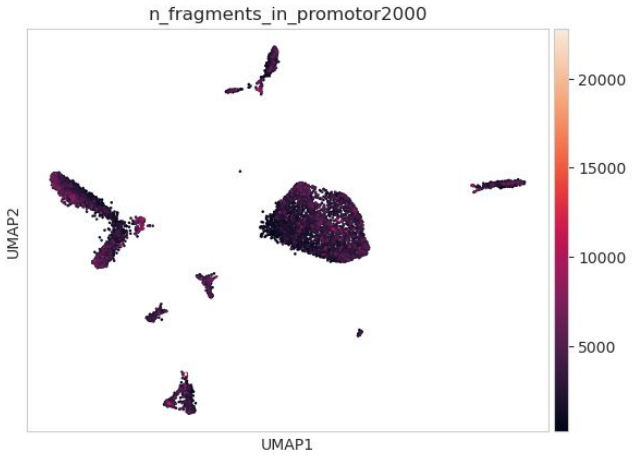


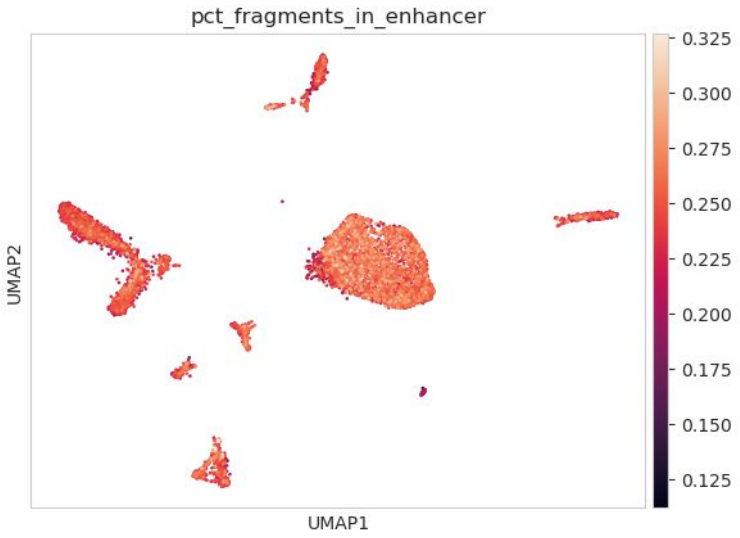
UMAP

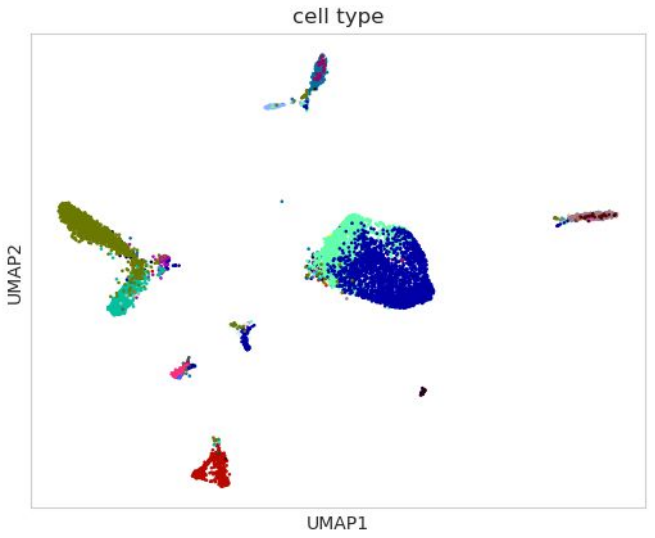


Dimensionreduction - Colon

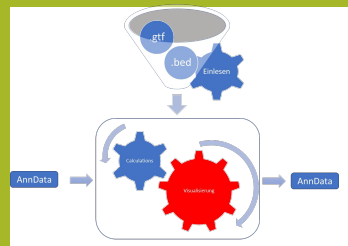




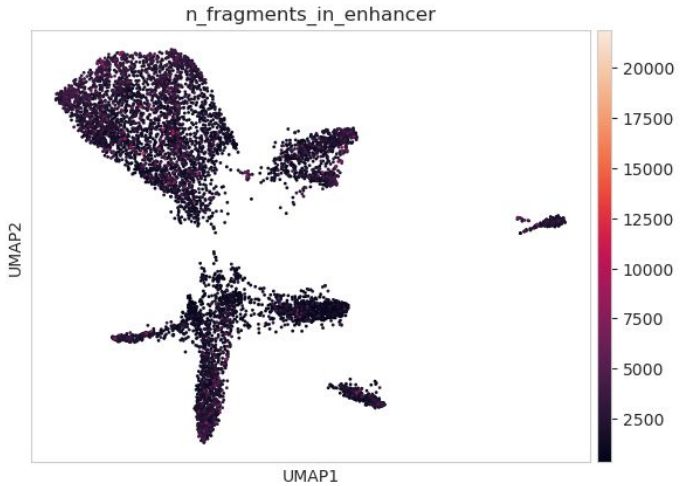


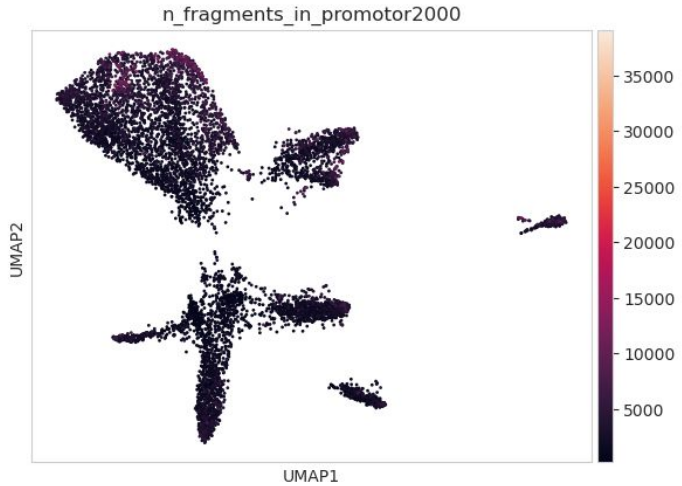


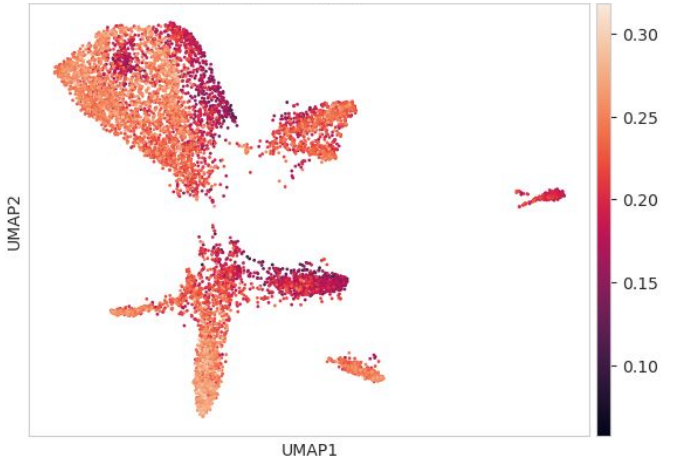
- Adipocyte
- CNS,Enteric Neuron
- Cardiac Fibroblasts
- Cardiac Pericyte 2
- Cardiac Pericyte 3
- Endothelial Cell (General) 1
- Endothelial Cell (General) 2
- Endothelial Cell (Myocardial)
- Fibroblast (Epithelial)
- Fibroblast (Gastrointestinal)
- Fibroblast (General)
- Fibroblast (Peripheral Nerve)
- Luteal Cell (Ovarian)
- Lymphatic Endothelial Cell
- Macrophage (General)
- Macrophage (General,Alveolar)
- Mammary Luminal Epithelial Cell 2
- Mast Cell
- Memory B Cell
- Naive T cell
- Pericyte (Esophageal Muscularis)
- Pericyte (General) 1
- Pericyte (General) 2
- Pericyte (General) 3
- Schwann Cell (General)
- Smooth Muscle (Colon) 1
- Smooth Muscle (Colon) 2
- Smooth Muscle (Esophageal Muscularis) 1
- Smooth Muscle (Esophageal Muscularis) 2
- Smooth Muscle (Esophageal Muscularis) 3
- Smooth Muscle (GE Junction)
- Smooth Muscle (General Gastrointestinal)
- Smooth Muscle (General)
- Smooth Muscle (Uterine)
- Smooth Muscle (Vaginal)
- T Lymphocyte 1 (CD8+)
- Vascular Smooth Muscle 2

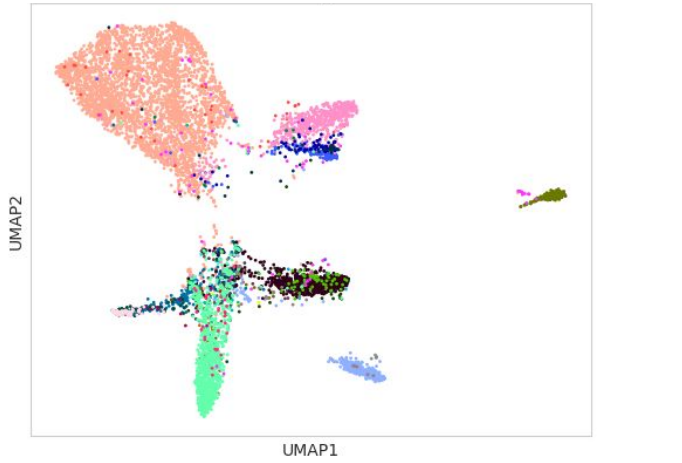


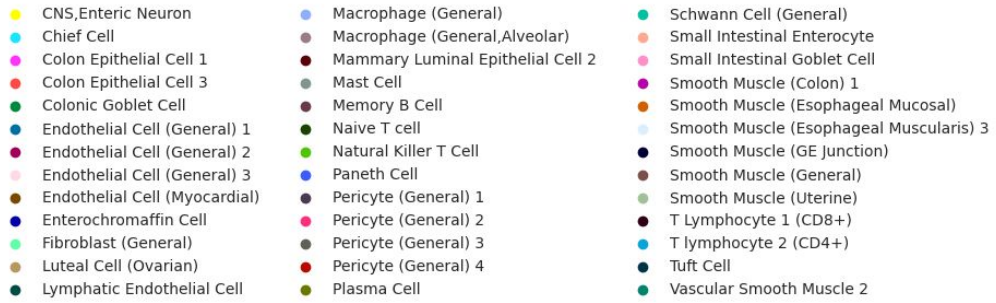
Dimensionreduction - Intestinal











Einordnung

Ausblick

Was ist weiterhin geplant ?

- Verbesserung des Scores
- Einführung von Kategorien
- Einbindung von TSS Plots in den Workflow
- Schnittstelle zwischen WP1/WP2
  - Was ist mit dem X passiert?
- Plotting Tools automatisieren
- Plotting optimieren

Welche Fragen haben Sie noch ?