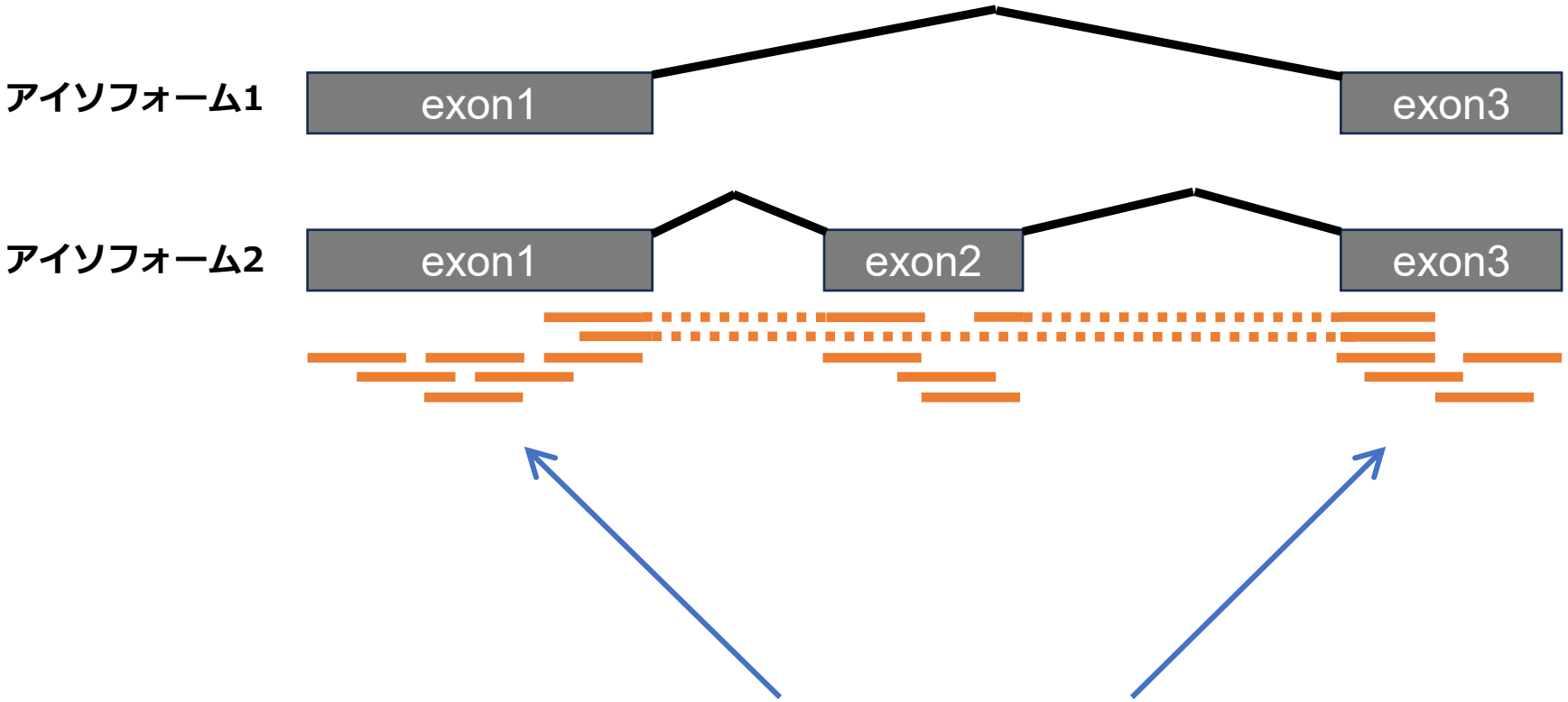
ロングリードを使用したRNA-seq解析

フィルジェン株式会社　バイオインフォマティクス部
（ support@filgen.jp ）

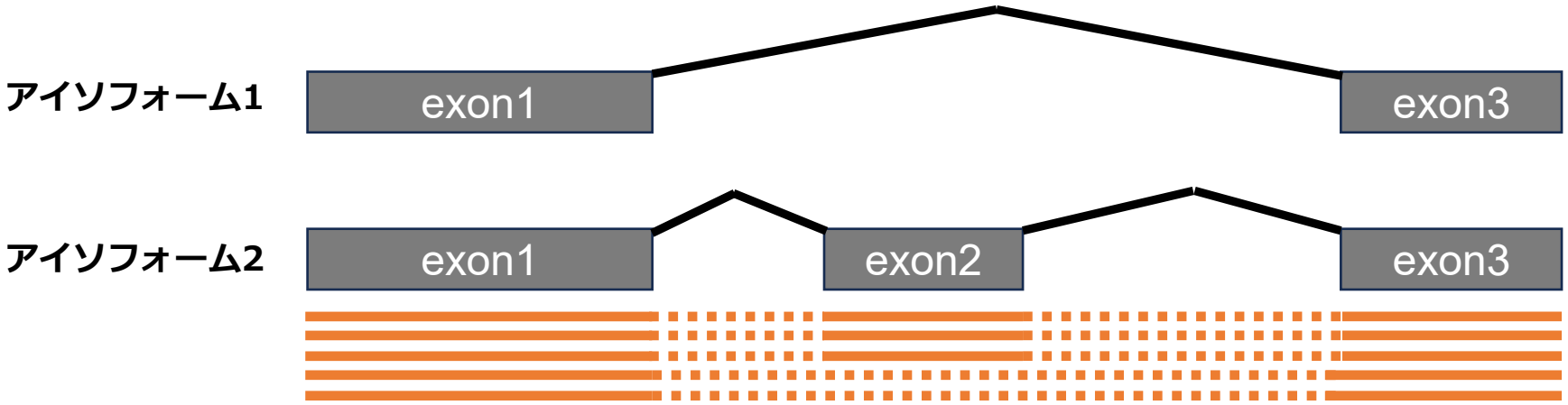
ロングリードを使用したRNA-seqのメリット



ショートリードの場合、多くのリードがエクソン内で完結するため、
どっちのアイソフォームに由来するリードか判別不能

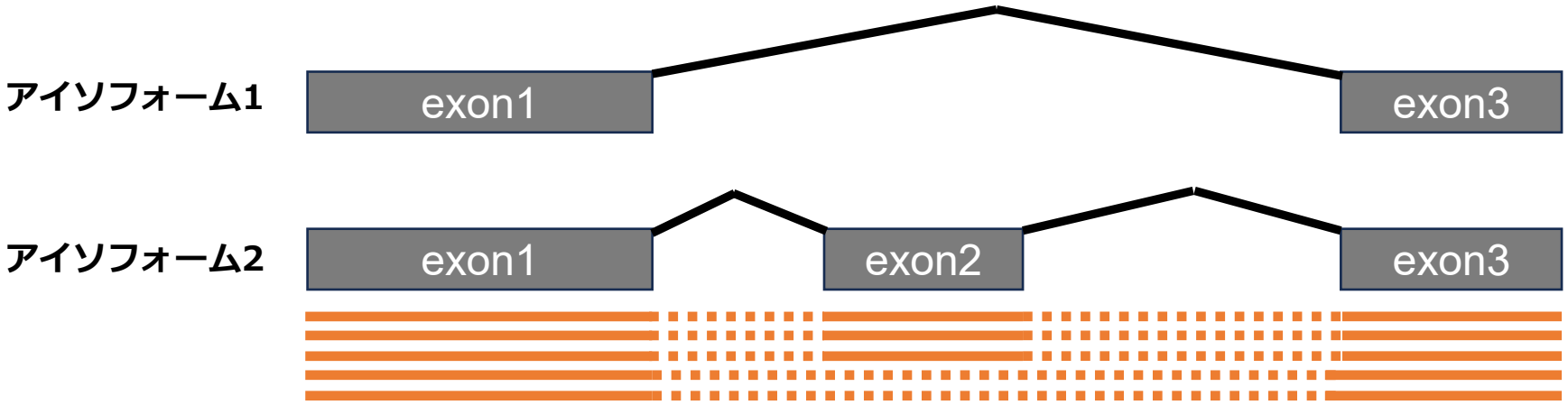
→<u>アイソフォームごとに定量することが困難</u>

ロングリードを使用したRNA-seqのメリット



ロングリードの場合、ほとんどのリードが転写産物全長の配列をもつため、
転写産物ごとの定量が容易

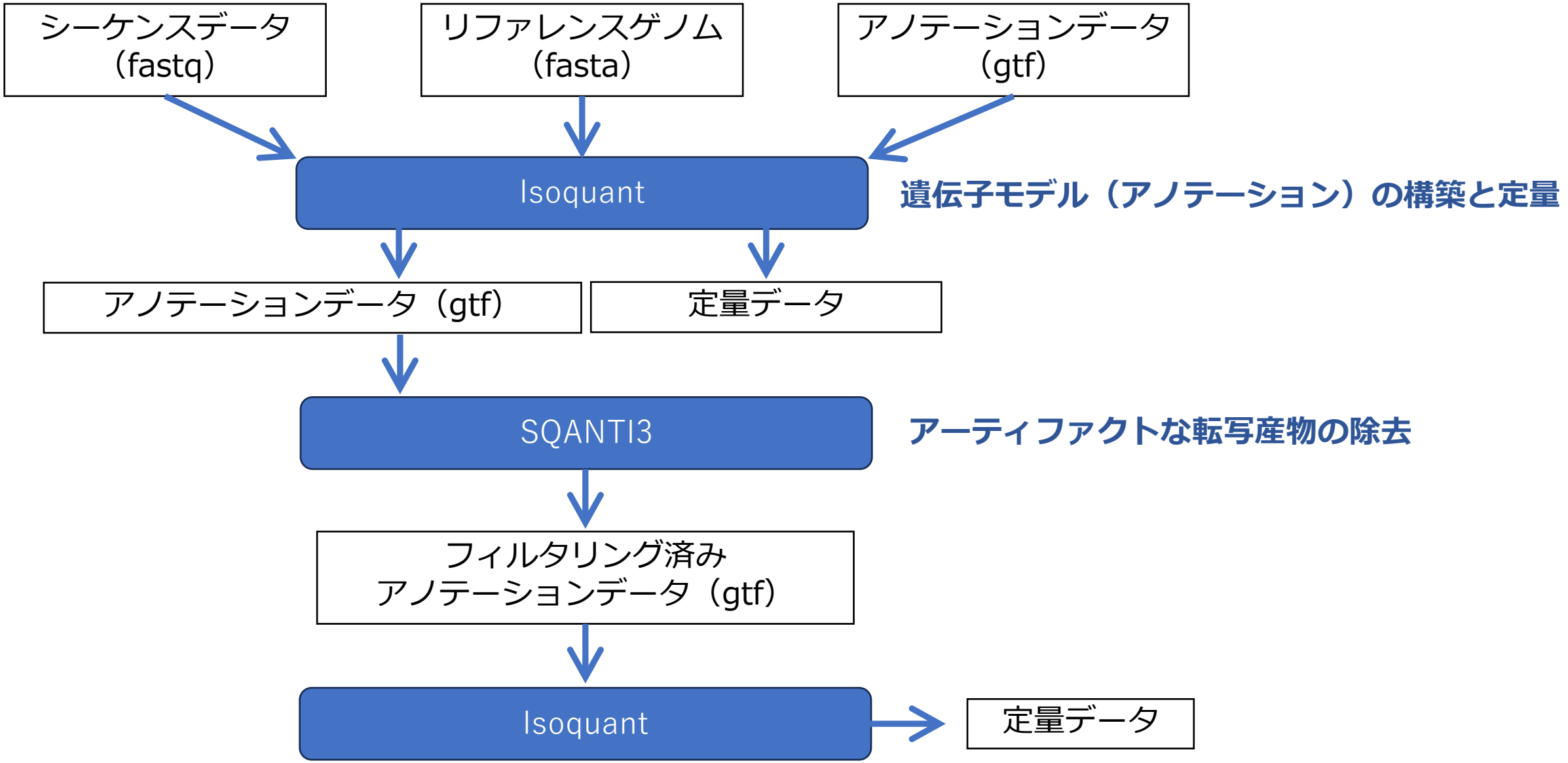
ロングリードを使用したRNA-seqのメリット



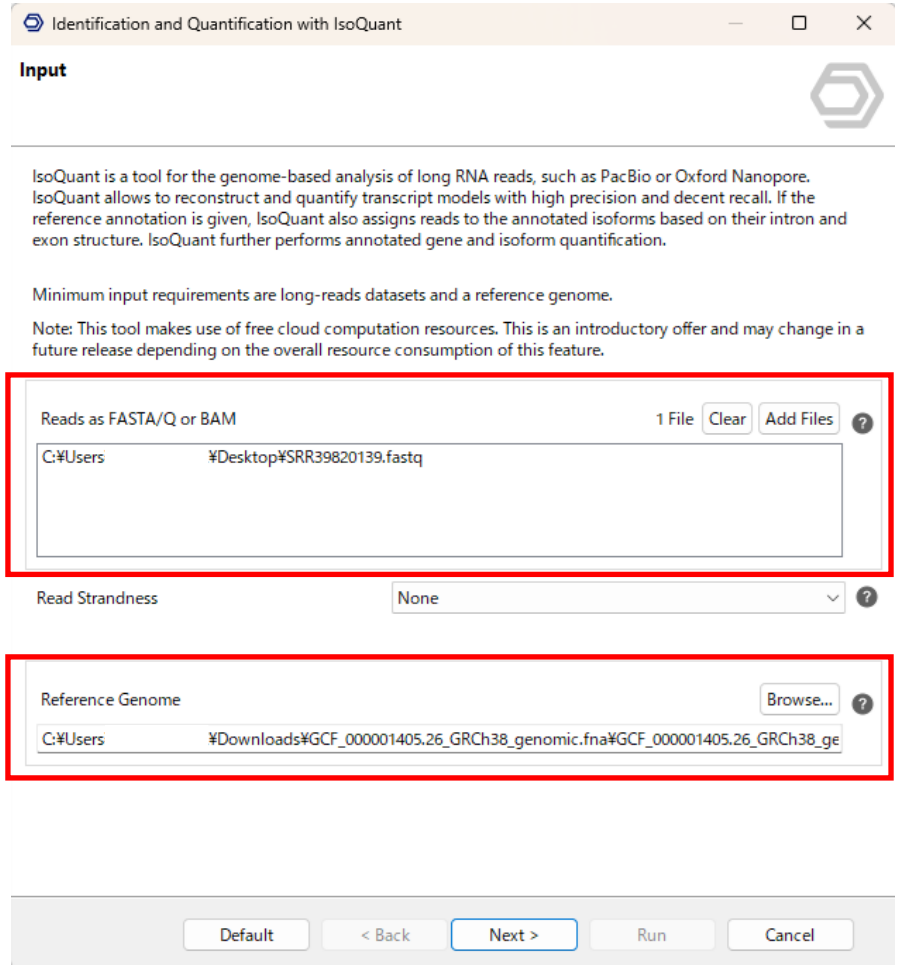
ロングリードの場合、ほとんどのリードが転写産物全長の配列をもつため、
転写産物ごとの定量や新規転写産物の発見に有利

ただし、ロングリードではショートリードに比べてエラーが多いため、
アーティファクトな転写産物が生じる場合がある。

ロングリードを使用したRNA-seqのパイプライン



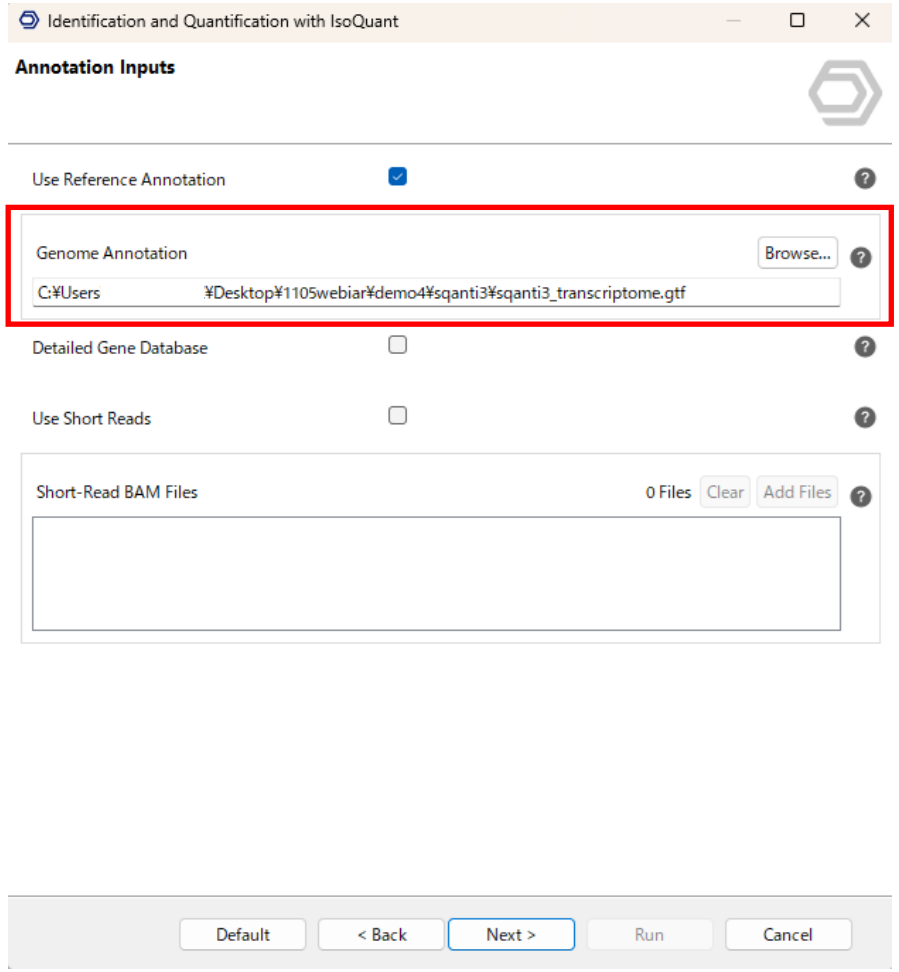
IsoQuant



シーケンスデータ（fastq）

リファレンスゲノム（fasta）

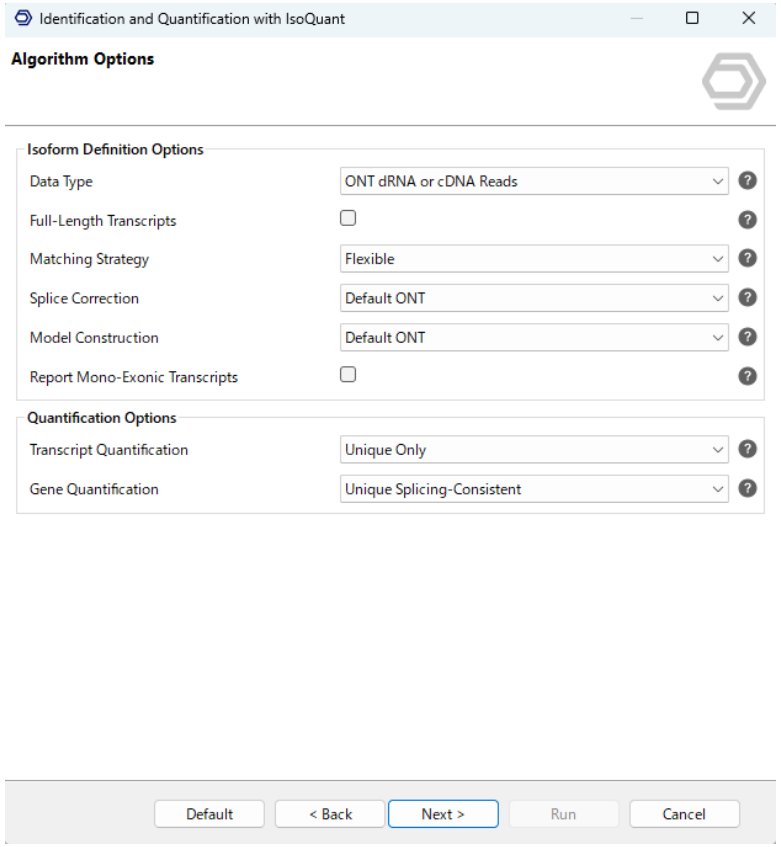
IsoQuant



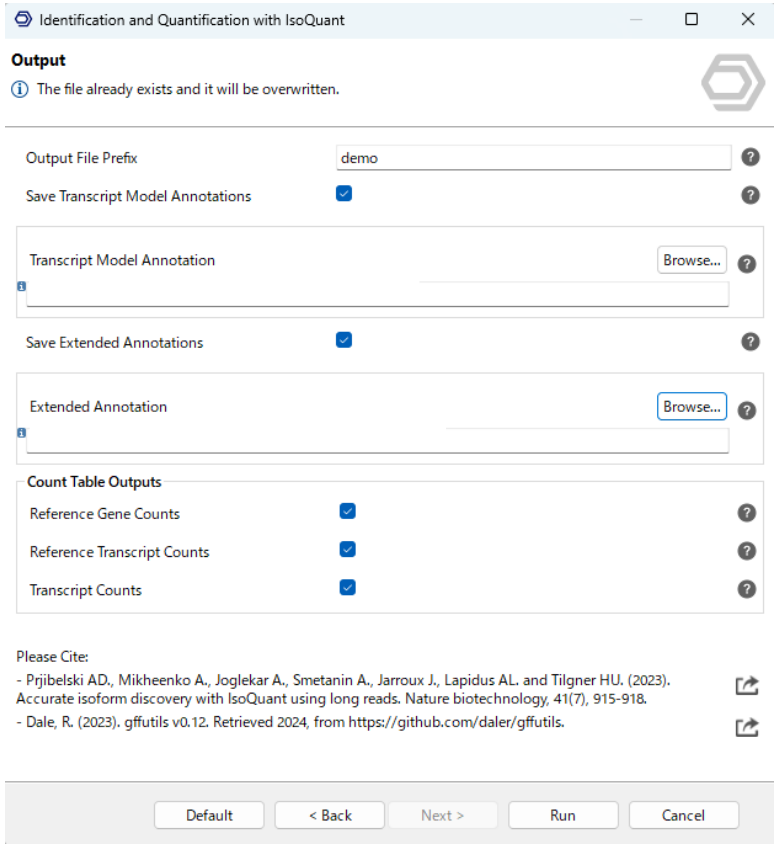
アノテーションデータ（gtf）

ショートリードのマッピングデータ（オプション）

IsoQuant



シーケンスプラットフォームや検出感度を指定



出力するデータおよび出力先を指定

IsoQuantの出力

demo_transcript_model_counts.box

demo_reference_transcript_counts.box

Feature	count
NM_000016.5	2
NM_000029.3	9
NM_000130.4	1
NM_000143.3	3
NM_000157.3	2
NM_000191.2	11
NM_000228.2	13
NM_000302.3	7
NM_000310.3	10
NM_000374.4	1
NM_000396.3	2
NM_000642.2	2
NM_000701.7	37
NM_000702.3	1
NM_000760.3	1
NM_000831.3	1
NM_000848.3	17

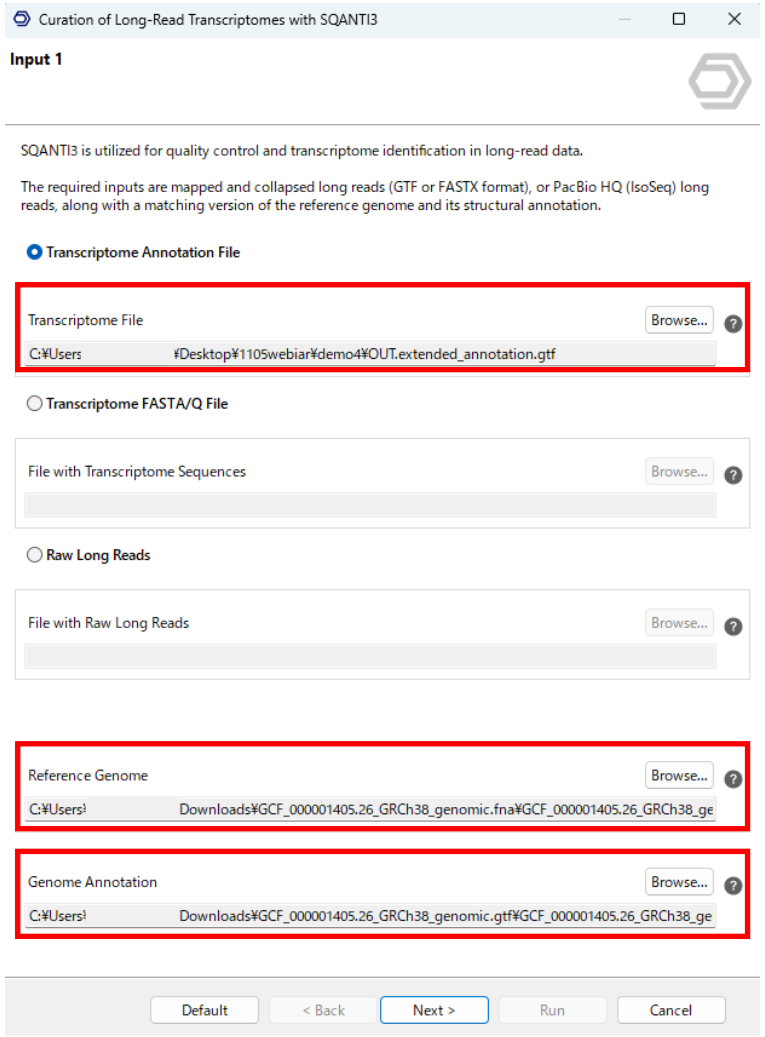
スプライシングバリアントごとのカウントデータ
transcript_model_counts: IsoQuantで提案された転写産物モデルに準拠
reference_transcript_counts: リファレンスで提供された転写産物モデルに準拠

OUT.extended_annotation.gtf

OUT.transcript_models.gtf

アノテーションデータ
extended_annotation: IsoQuantで提案された転写産物モデルとリファレンスで提供されたモデル
Transcript_models: : IsoQuantで提案された転写産物モデル

SQANTI3

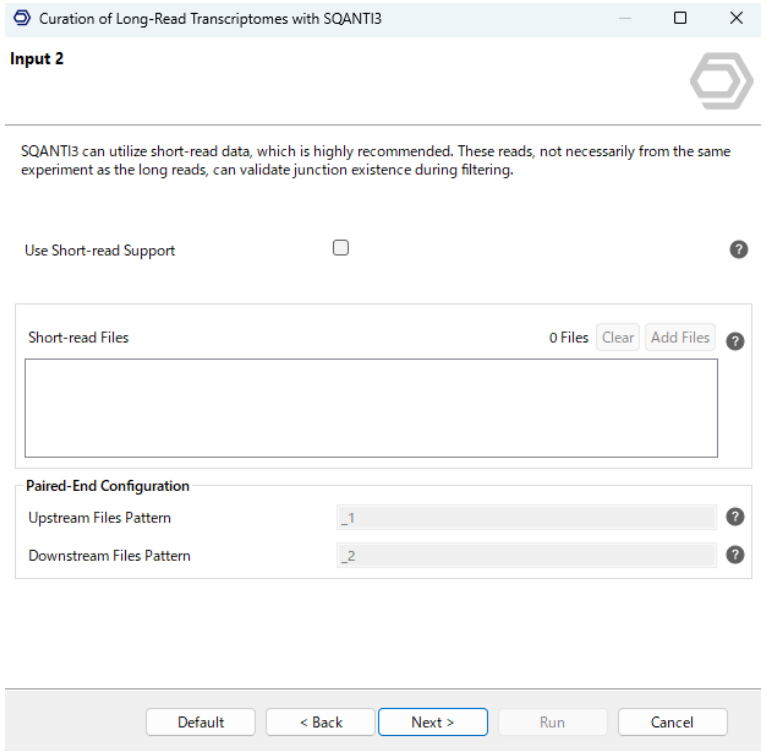


更新済みアノテーションデータ（gtf）

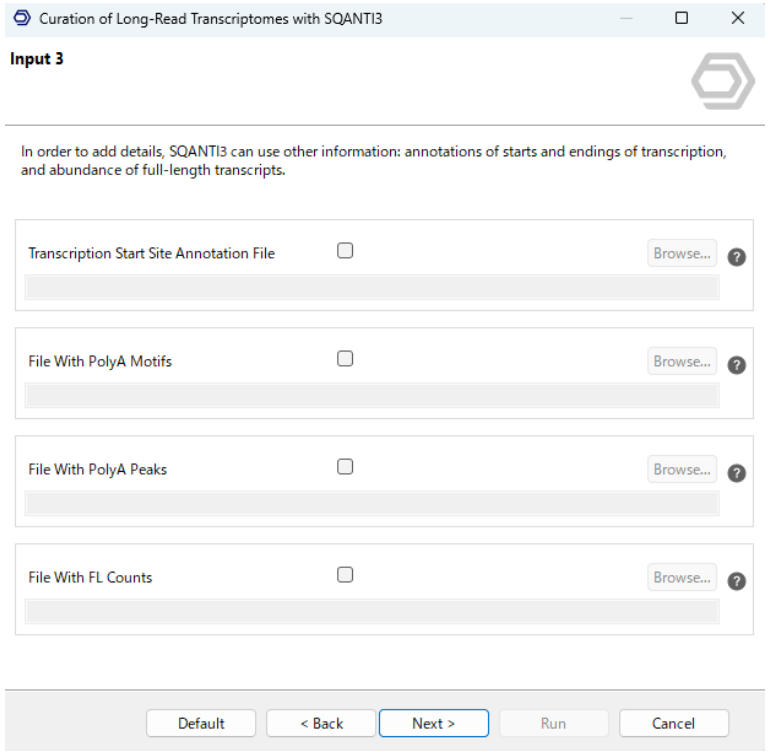
リファレンスゲノム（fasta）

リファレンスアノテーション（gtf）

SQANTI3

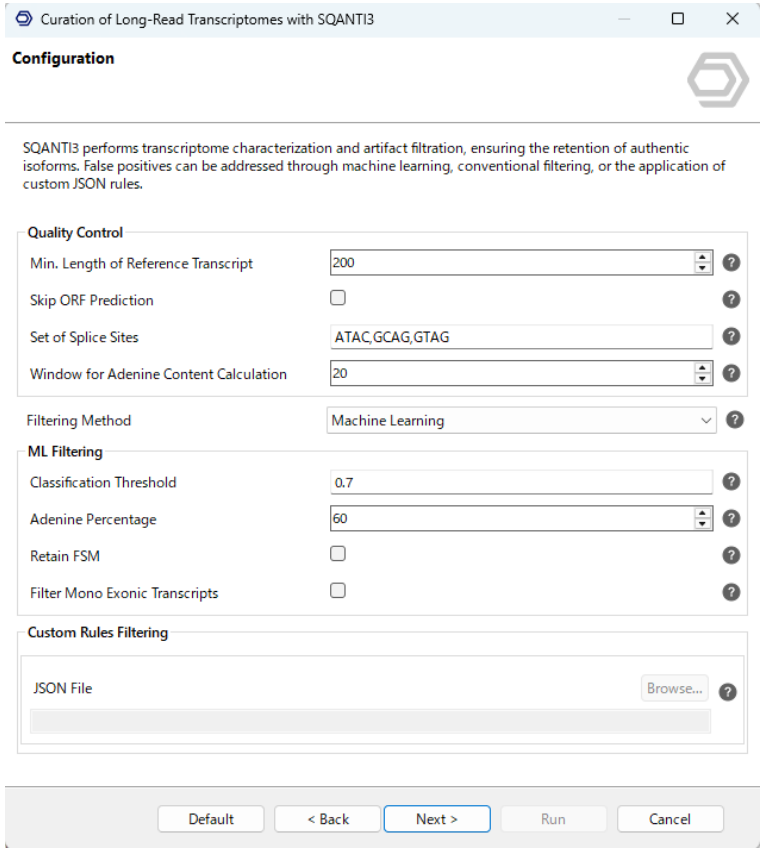


ショートリードのマッピングデータを指定（オプション）

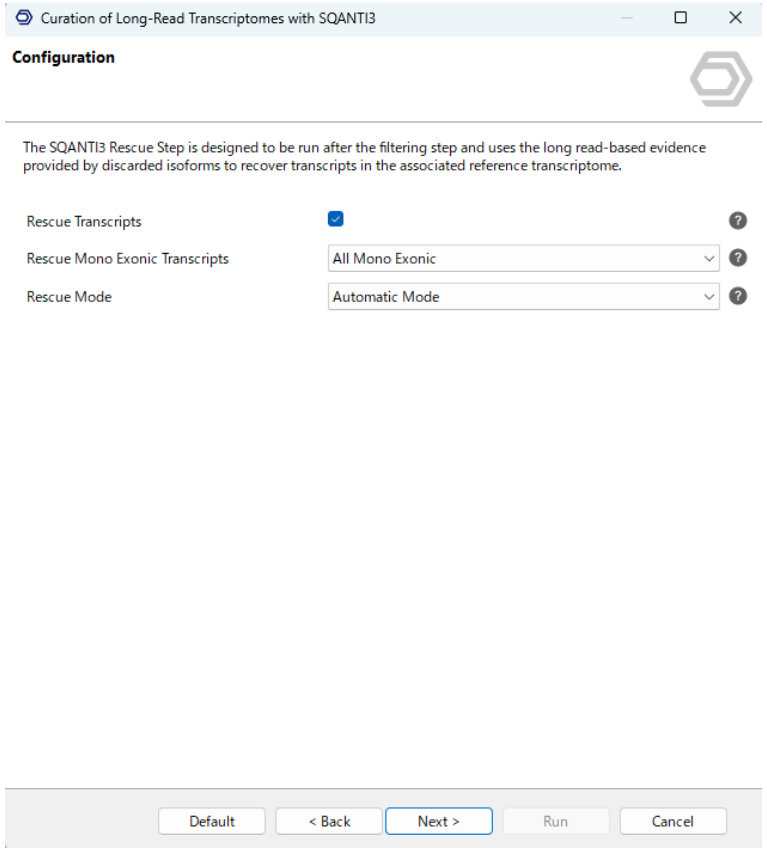


転写開始位置データなどの補足データを指定（オプション）

SQANTI3



フィルタリングの閾値等を設定



転写産物のレスキューを行うかどうかを指定

SQANTI3の出力

sqanti3_report.box

sqanti3_results.box

転写産物の分類とレポート

results

Category	Subcategory	Isoform	Chr	Length	Exons	Gene ID	Transcript ID	Gene ...	Differ...	Differ...	RT swi...	All Ca...	Codin...	Predic...
Intergenic	Multi-exon	transcript705.N...	NC_000006.12	1092	2	novelGene_2	novel	.	.	.	False	Yes	coding	False
Intergenic	Multi-exon	transcript210.N...	NC_000001.11	348	2	novelGene_1	novel	.	.	.	False	Yes	non_cod...	-
Novel Not In Catalog	1+ donor / acceptor	transcript747.N...	NC_000019.10	495	4	C19orf33	novel	.	.	.	False	Yes	non_cod...	-
Novel Not In Catalog	1+ donor / acceptor	transcript70.NC_...	NC_000019.10	609	4	RPS15	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript623.N...	NC_000016.10	223	2	MT1G	novel	.	.	.	False	Yes	non_cod...	-
Novel Not In Catalog	1+ donor / acceptor	transcript378.N...	NC_000016.10	751	4	ZG16	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript489.N...	NC_000012.12	499	4	RPL41	novel	.	.	.	False	Yes	non_cod...	-
Novel Not In Catalog	1+ donor / acceptor	transcript809.N...	NC_000011.10	526	6	RPS3	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript764.N...	NC_000011.10	1359	7	STARD10	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript92.NC_...	NC_000011.10	1087	4	IGF2	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript984.N...	NC_000007.14	1156	7	TMEM176A	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript256.N...	NC_000006.12	593	3	TOMM6	novel	.	.	.	False	Yes	non_cod...	-
Novel Not In Catalog	1+ donor / acceptor	transcript1642....	NC_000001.11	595	12	GAS5	novel	.	.	.	False	Yes	non_cod...	-
Novel Not In Catalog	1+ donor / acceptor	transcript1367....	NC_000001.11	459	3	S100A6	novel	.	.	.	False	Yes	coding	False
Novel Not In Catalog	1+ donor / acceptor	transcript1358....	NC_000001.11	580	4	S100A6	novel	.	.	.	False	Yes	coding	False
Novel In Catalog	Intron retention	transcript616.N...	NC_000023.11	848	6	RPL10	novel	.	.	.	False	Yes	coding	False
Novel In Catalog	Intron retention	transcript268.N...	NC_000023.11	1844	6	IGBP1	novel	.	.	.	False	Yes	coding	False

Type of Isoform	Number before filtering	Number after filtering
Full Splice Match	95268 (99.88%)	93569 (99.89%)
Incomplete Splice Match	80 (0.08%)	72 (0.08%)
Novel In Catalog	11 (0.01%)	11 (0.01%)
Novel Not In Catalog	20 (0.02%)	13 (0.01%)
Genic Genomic	0 (0.00%)	0 (0.00%)
Antisense	1 (0.00%)	0 (0.00%)
Fusion	0 (0.00%)	0 (0.00%)
Intergenic	4 (0.00%)	4 (0.00%)
Genic Intron	0 (0.00%)	0 (0.00%)
Total	95384 (100.00%)	93669 (100.00%)

report

Full Splice Match:
リファレンスとの完全一致

Novel In Catalog:
既知エキソンの新規組み合わせ

Novel Not In Catalog:
新規エキソン

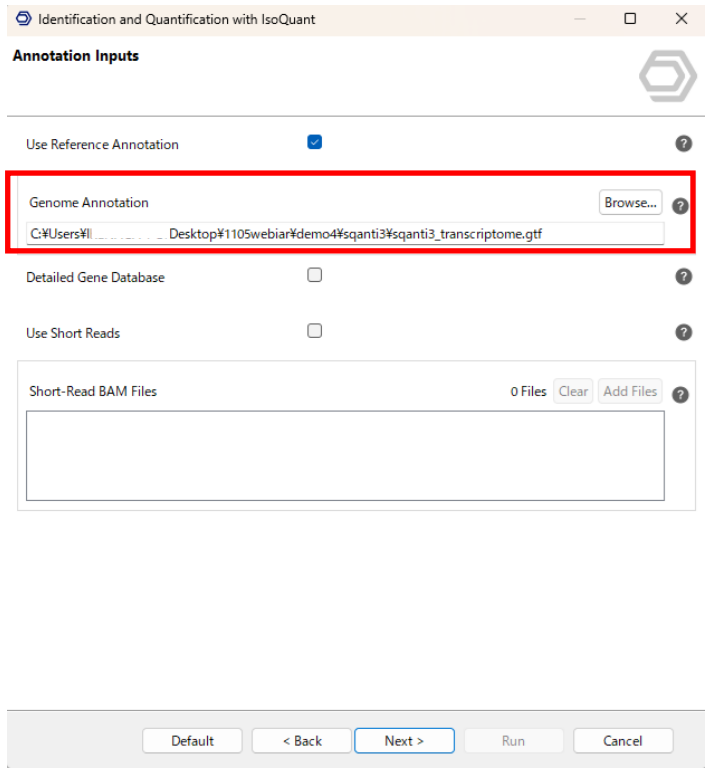
Intergenic:
遺伝子間領域

など

IsoQuant2（2回目）



フィルタリング済みのアノテーションデータ



SQANTI3で出力されたアノテーションデータを使用し、
再度IsoQuantを実行

お問い合わせ先：フィルジェン株式会社

TEL: 052-624-4388 (9:00～17：00)

FAX: 052-624-4389

E-mail:　support@filgen.jp