Aplikační nota

Precizní detekce AML MRD: Integrace technologie µCaler™ s platformou Variant Analyzer

Tomáš Hron[1], Tereza Valterová[2], Lenka Žejšková[2], Markéta Kalinová[2]

[1] Analýza dat a vzdělávání v genomice a sekvenačních technologiích, e-mail: tomas.hron@bioprojects.cz, telefon: +420 605 325 350

[2] Centrální laboratoře, Laboratoř genomické medicíny, Fakultní nemocnice Královské Vinohrady, Praha, Česká republika

Abstrakt

Diagnostika akutní myeloidní leukémie (AML) a monitorování minimální reziduální nemoci (MRD) vyžadují detekci variant při extrémně nízkých frekvencích alel. Tato studie hodnotí komplexní řešení kombinující přípravu knihoven Nanodigmbio µCaler™ pro AML MRD s analyzační platformou Variant Analyzer (VARAN).

Deduplikace a použití konsensuálních čtení pro korekci artefaktů pomocí UMI a optimalizace specifických parametrů umožňuje platformě VARAN detekci variant s extrémně nízkou frekvencí. Poskytujeme dva různé analytické postupy, které vyvažují analytickou senzitivitu a zároveň udržují nízký počet falešně pozitivních výsledků. Náš přístup je schopen efektivně detekovat varianty až do 0,05 % VAF, což se blíží technickým limitům přístupu NGS. Díky tomu je toto řešení obzvláště vhodné pro monitorování MRD.

Řešení VARAN pro diagnostiku AML a MRD

- Kombinace analytické platformy VARAN a knihoven Nanodigmbio µCaler™ umožňuje detekci variant s extrémně nízkou frekvencí, které jsou vhodné pro monitorování minimální reziduální nemoci (Minimal Residual Disease – MRD).

Maximální detekční limit a citlivost

- Poskytujeme dva různé analytické postupy (workflow – wf), které vyvažují citlivost a specificitu. Náš přístup je schopný efektivně detekovat varianty až do 0,05% VAF, což se blíží technickým limitům přístupu NGS.

Doporučené experimentální nastavení

- **Citlivé (>0,05% VAF):** Sensitive wf, 5 milionů párů čtení na vzorek
- **Citlivé a úsporné (>0,05% VAF, kompromitovaná specificita):** Sensitive wf, 1milion párů čtení na vzorek
- **Vyvážené (>0,45% VAF, vysoká specificita):** Balanced wf, 1 milion párů čtení na vzorek
- **Vyvážené a úsporné (>0,5% VAF, vysoká specificita):** Balanced wf, 500 000 párů čtení na vzorek
-

Úvod

Akutní myeloidní leukémie (AML) je hematologické onemocnění charakterizované rychlým růstem nádorových buněk v kostní dřeni. Minimální reziduální nemoc (MRD) označuje malý počet rakovinných buněk, které zůstávají v těle pacienta během nebo po léčbě a slouží jako klíčový prognostický indikátor pro sledování relapsu a reakce na léčbu. Detekce těchto vzácných klonů vyžaduje analytické přístupy s vysokou analytickou senzitivitou. Kombinace technologie přípravy NGS knihoven od firmy Nanodigmbio a platformy VARiant ANalyzer (VARAN) umožňuje detekci variant s extrémně nízkou frekvencí, která je nezbytná pro hodnocení MRD.

Panel µCaler™ AML v1.0 od společnosti Nanodigmbio je navržen specificky pro náročné požadavky monitorování akutní myeloidní leukémie u dospělých. Panel cílí na 32 klinicky významných genů pokrývající oblast o velikosti ~42,5 Kb. Je navržen pro cílenou detekci více typů mutací, včetně substitucí, malých inzercí/delecí (indely) a genových fúzí. Metoda nabohacení vybraných genomických oblastí pomocí hybridizace se specifickými probami je speciálně vyladěna tak, aby zachytila vzácné maligní klony, i když představují pouze zlomek celkové buněčné populace.

VARAN je sofistikovaná webová platforma poskytovaná společností Altium International, která je navržena tak, aby transformovala surová sekvenační data na využitelné klinické a výzkumné poznatky, s důrazem na bezpečnost a přizpůsobitelnost. Všechny výpočty jsou prováděny na výkonných a bezpečných cloudových serverech umístěných v České republice, což zajišťuje ochranu citlivých klinických dat. VARAN využívá sadu nástrojů Sentieon® pro identifikaci variant, která je uznávána jako světový lídr v přesnosti detekce. Analytické postupy lze přizpůsobit konkrétním sekvenačním knihovnám, platformám a diagnostickým potřebám. Pro aplikace detekce somatických variant podporuje VARAN zpracování UMI sekvencí (Unique Molecular Identifiers) za účelem korekce sekvenačních artefaktů a detekce nízkofrekvenčních variant.

Materiály a metody

Vzorky

Referenční vzorek (testovací vzorek) byl zředěn do Buffy Coat pool (10 germinálních vzorků) k vytvoření ředící řady vzorků: Vzorek 1 (1:9 ředění testovacího vzorku), Vzorek 2 (1:99 ředění), Vzorek 3 (1:999 ředění).

Příprava knihovny a sekvenace

Knihovny byly vytvořeny pomocí µCaler™ AML MRD a sekvenovány na sekvenátoru DNBSEQ-T1+ (MGI) s chemií 2x150 bp. V tabulce 1 naleznete seznam genů a jejich pokrytí panelem µCaler AML Panel v1.0. Pro vyhodnocení efektu sekvenačního pokrytí na výsledek detekce byla data subsamplována od 50 milionů až po 10 tisíc párů čtení, čímž vzniklo 8 paralelních datových sad (50M, 10M, 5M, 1M, 500K, 100K, 50K, 10K). Kvalita sekvenování dosáhla Q30 pro více než 96 % a Q40 pro více než 90 % bází.

Analýza dat v platformě VARAN

Data byla zpracována pomocí sady nástrojů Sentieon integrované do platformy VARAN. Pro zmírnění sekvenačních artefaktů a zlepšení specificity při nízkém VAF byly použity UMI sekvence k vytvoření duplexních konsensuálních čtení. Byly porovnány čtyři workflows: i) Bez UMI (standardní somatická detekce bez deduplikace a korekce pomocí UMI), ii) Normální (standardní somatická detekce s použitím čtení opravených pomocí UMI), iii) Vyváženě citlivá – Balanced (optimalizovaná pro nízké VAF s vyváženou specificitou) a iv) Vysoce citlivá – Sensitive (optimalizovaná pro extrémně nízké VAF).

Definice očekávaných variant

Pro vyhodnocení efektivity řešení byla vytvořena sada očekávaných variant. Genetické varianty (SNV a indels) Buffy Coat poolu a testovacího vzorku byly stanoveny pomocí nezávislého NGS panelu (SureSelect target enrichment HS XT) a nástroje Varsome Clinical. Celkem 45 variant s VAF > 3 %, nacházející se v regionech panelu µCaler™ AML, bylo použito jako referenční sada.

Tabulka 1. Seznam genů pokrytých panelem µCaler AML Panel v1.0.

ASXL1 Exon 12,13	*BRINP3* Exon 3,8	*CBL* Exon 8,9	*CEBPA** Full CDS	*DNMT3A* Exon 8-23	*EZH2* Exon 4-6,8,13-20	*FLT3* Exon 14,15,20	*GATA2* Exon 3-6	*HNRNPK* Exon 4-6,10,12,15,16	*IDH1* Exon 4	*IDH2* Exon 4
JAK2 Exon 14	*KIT* Exon 8,17	*KMT2A*† Intron 8-10	*KRAS* Exon 2,3	*MYH11*† Intron 32	*NPM1* Exon 10,11	*NRAS* Exon 2,3	*PHF6** Full CDS	*PTEN* Exon 5,7	*PTPN11* Exon 3,13	*RAD21** Full CDS
*RUNX1** Full CDS	*SF3B1* Exon 14,15	*SMC1A* Exon 2,9,11,13,15-17,22		*SMC3* Exon 9,10,13,19,24,25,27		*SRSF2* Exon 1	*STAG2* Exon 5,6,8-10,15,17,19,20,25,27-31		*TET2** Full CDS	*TP53** Full CDS
U2AF1 Exon 2	*WT1* Exon 7-10									

* *Označuje, pokrytí celé protein-kódující oblasti genu.*
† *Označuje, pokrytí intronické oblasti pro detekci genové fúze.*

Pro zředěné vzorky (Vzorek 1, 2 a 3) byla očekávaná VAF pro každou variantu vypočítána jako vážený součet VAF z Buffy Coat a Testovacího vzorku, zohledňující jednotlivá ředění. Falešně pozitivní varianty byly predikovány pro VAF >0,1 % diferenciací mezi ředěními vzorků. Konkrétně falešně pozitivní varianty byly definovány jako varianty přítomné ve vzorku 3 s VAF > 0,1 % a zároveň nepřítomné ve vzorku 1 (VAF limit 1 %).

Výpočet citlivosti a FPR

Citlivost byla vypočítána na základě celkového souboru všech kombinací variant a vzorků ve vzorku 1, vzorku 2 a vzorku 3. V tomto modelu se jedna varianta počítá vícekrát, pokud je nezávisle identifikována napříč různými vzorky a úrovněmi zředění, což poskytuje širší pohled na detekční limity. Abychom získali standardizovanou metriku pro specificitu, byla vypočítána FPR (False Positive Rate) jako počet falešně pozitivních variant na kilobázi (kbp) cílové oblasti panelu.

Výsledky a diskuse

Zvýšení specificity pomocí UMI konsensu

Detekce nízkofrekvenčních SNVs při monitorování MRD je primárně omezena technickým šumem NGS technologie. I když moderní sekvenační platformy dosahují relativně vysoké kvality – sekvenační chybovost v řádech 0,01 % (Q40) – hlavním úskalím zůstává PCR reakce během zpracování knihovny. V důsledku je tedy pozorován výskyt arteficielních variant s VAF dosahujícím téměř 1 %.

Ačkoli analýza dat umožňuje detekovat varianty s nízkou frekvencí, úskalím bývá často právě odlišení těchto chyb. Pro překonání tohoto problému VARAN využívá tvorbu duplexního konsensuálního čtení pomocí UMI sekvencí, což efektivně snižuje šum v datech a umožňuje rozlišit skutečné biologické varianty od technických artefaktů.

Jak ukazuje obrázek 1, workflow s deduplikací a korekcí chyb pomocí UMI (UMI cons.) a normální workflow (bez UMI) vykazují podobný výsledek při detekci variant s VAF nad 1,0 %. Jakmile se však detekční limit posune pod tuto hranici, standardní workflow vykazuje prudký nárůst počtu identifikovaných variant. Tato divergence přímo odpovídá očekávané kumulativní chybovosti PCR amplifikace a sekvenační chemie.

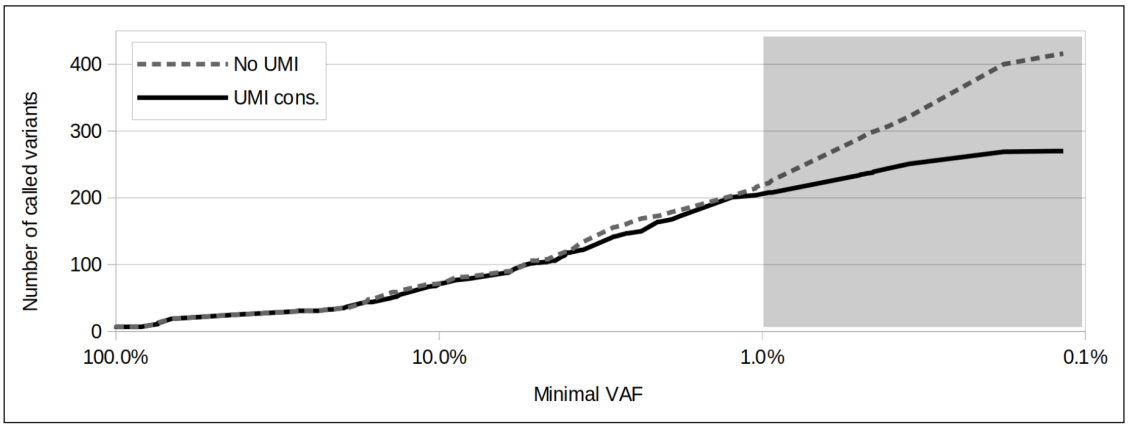
Analýza s UMI zachovává vysokou analytickou citlivost a zároveň drasticky snižuje míru falešně pozitivních výsledků (FPR). To ukazuje, že deduplikace a korekce s použitím UMI konsenzu je nezbytnou součástí spolehlivé detekce variant v citlivém režimu, nezbytným pro identifikaci vzácných klonů.

Citlivost analýzy pomocí nástroje VARAN

Pro naplnění různých požadavků klinického výzkumu a monitorování MRD byly v platformě VARAN vyvinuty dvě specializované workflows: **Balanced** a **Sensitive**. Tyto workflows vyvažují detekční limity vůči míře falešně pozitivních výsledků (FPR). Výsledky byly posuzovány pro vzorky se sekvenačním pokrytím 1 milion párů čtení na vzorek, což odpovídá sekvenančí hloubce přibližně 6 400x. Rozdíl mezi jednotlivými workflows je znázorněn na obrázku 2.

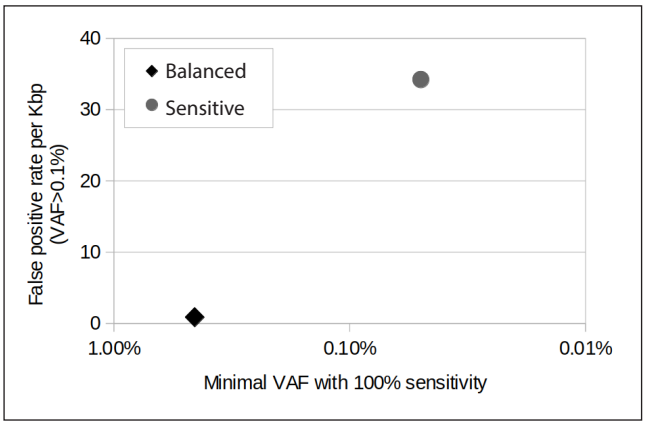
Workflow Balanced je vyvážena, aby poskytla vysokou citlivost s nízkým počtem falešně pozitivních výsledků a doporučuje se pro většinu případů. Poskytuje rozlišení 0,45% VAF při zachování nízké míry falešných pozitivních výsledků (méně než 1 falešně pozitivní výsledek na kbp).

Workflow Sensitive je optimalizována pro detekci extrémně nízké VAF při monitorování MRD. Dosahuje



Obrázek 1. Porovnání analýzy s a bez korekce pomocí UMI. *Graf zobrazuje počet variant detekovaných v závislosti na minimální detekované VAF pro všechny testované vzorky se sekvenačním pokrytím 1 milion párů čtení na vzorek. Přerušovaná čára představuje normální workflow (bez UMI), zatímco plná čára ukazuje výsledky po deduplikaci a korekci chyb pomocí UMI (UMI cons.). Šedá plocha zvýrazňuje kritickou oblast VAF, kde se technický šum (chyby PCR/sekvenování) stává dominantním faktorem ve workflows bez UMI (No UMI).*

rozlišení 0,05 % VAF. Tato zvýšená citlivost je dosažená na úkor vyššího počtu artefaktů (přibližně 34 falešných pozitivních výsledků na kbp), což je v kontextu detekce konkrétních MRD variant obecně akceptovatelné.



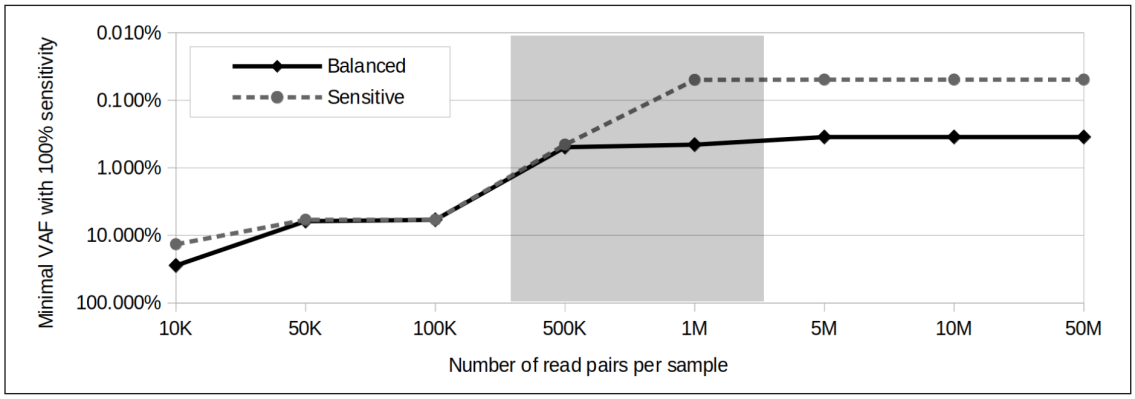
Obrázek 2. Citlivost a FPR workflows v platformě VARAN. *Graf zobrazuje minimální hodnotou VAF potřebnou k dosažení 100% citlivosti a výslednou mírou falešně pozitivních výsledků (FPR) na kilobázi cílového panelu. Datové body představují workflow Balanced a workflow Sensitive při sekvenančím pokrytí 1 milionu párů čtení na vzorek.*

Vztah mezi hloubkou sekvenačního pokrytí a analytickou citlivostí

Abychom určili optimální sekvenační pokrytí pro monitorování MRD, provedli jsme analýzu vzorků s pokrytím od 10 tisíc do 50 milionů párů čtení na vzorek.

Jak je znázorněno na obrázku 3, citlivost se výrazně zlepšuje s rostoucí hloubkou sekvenování z 10 tisíc na 500 tisíc párů čtení. Mezi 500 tisíc a 1 milionem párů čtení však dochází k ustálení. Zvýšení hloubky sekvenování nad 1 milion párů čtení vede k významnému zvýšení míry PCR duplikátů, které dosahuje 96,56 % při 50 milionech párů čtení (tabulka 2). V tomto bodě již další sekvenační kapacita ztrácí efekt, protože knihovna již jednoduše neobsahuje další unikátní fragmenty. Ačkoli jsou tyto duplikáty užitečné pro korekci chyb pomocí UMI deduplikace, nevedou ke zvýšení rozlišení nad prahovou hodnotu 0,05 % VAF.

Pro obě workflows proto doporučujeme 1 milion párů čtení na vzorek jako optimální výstup sekvenování, což odpovídá ~ 6 400x hrubému pokrytí. Dostatečná citlivost je zajištěna při 0,45 % VAF pro Balanced workflow a při 0,05 % VAF pro Sensitive workflow. U Balan-



Obrázek 3. Minimální VAF se 100% citlivostí při různých hloubkách sekvenování. *Graf ukazuje saturaci citlivosti pro obě workflows při přibližně 500 000–1 000 000 párech čtení. Šedá plocha označuje doporučený optimální rozsah sekvenování*

Tabulka 2. Teoretické vs. dosažené rozlišení metody (VAF) podle hloubky sekvenování.

Hloubka sekvenování [páry čtení]	Hrubé sekvenační pokrytí	PCR duplikáty [%]	Čisté sekvenační pokrytí	Teoretický VAF limit s 95% citlivostí*	Dosažená VAF (Sensitive)	Dosažená VAF (Balanced)
50M	320 031	96,56%	11 003	0,03%	0,05%	0,35%
10M	64 064	85,30%	9 401	0,03%	0,05%	0,35%
5M	32 032	75,16%	7 946	0,04%	0,05%	0,35%
1M	**6 406**	**40,97%**	**3 779**	**0,08%**	**0,05%**	**0,45%**
500K	3 203	26,70%	2 347	0,13%	0,45%	0,50%
100K	640	7,17%	594	0,50%	5,87%	5,87%
50K	320	3,69%	308	0,97%	5,87%	6,16%
10K	64	0,61%	64	4,70%	13,54%	27,88%

** Předpokládá se, že jedno čtení je dostačující pro detekci varianty a sekvenační pokrytí je rovnoměrné. Pro odhad sekvenační hloubky s 95% pravděpodobností pokrytí alespoň jedním čtením byla použita Poissonova distribuce.*

ced workflow může uživatel zvážit snížení výstupu sekvenování až na 500 tisíc párů čtení, přičemž lze očekávat podobný detekční limit.

Závěr

Kombinace knihovny µCaler™ a analýzy VARAN poskytuje robustní řešení pro monitorování MRD u pacientů s diagnozou AML. Pro optimální výkon doporučujeme sekvenační hloubku 1 milionu párů čtení na vzorek, což poskytuje nejlepší rovnováhu mezi citlivostí a sekvenančí kapacitou. Zatímco u workflow Balanced (detekce ~0,5 % VAF) může stačit 500 tisíc párů čtení, workflow Sensitive vyžaduje 1 milion párů čtení, aby dosáhla svého maximálního rozlišení 0,05 % VAF. Workflow Sensitive se tímto blíží technických limitů metodiky NGS založené na hybridizaci.

Reference

Altium International s.r.o je tvůrcem, vlastníkem a distributorem analyzační platformy VARiant ANalyzer (VARAN). Společnost také působí jako autorizovaný distributor analyzačního software, přístrojů a chemických kitů pro aplikace v molekulární biologii a genomice včetně portfolia Nanodigmbio.
altium.net/cz

VARiant ANalyzer (VARAN), je webová platforma zaměřená na hodnocení genetických variant v lidském genomu z dat sekvenování nové generace (NGS). Platforma umožňuje flexibilní zpracování dat, interpretaci výsledků a rozšiřování o pokročilé analýzy pro specifické potřeby uživatelů.
variantanalyzer.eu

Nanodigmbio Pte. Ltd., je výrobcem kompletní řady produktů pro cílené sekvenování NGS a poskytuje optimalizované sady pro aplikace v přesné medicíně a výzkumu.
nanodigmbio.com

Altium International s.r.o. 2026
www.altium.net/cz