



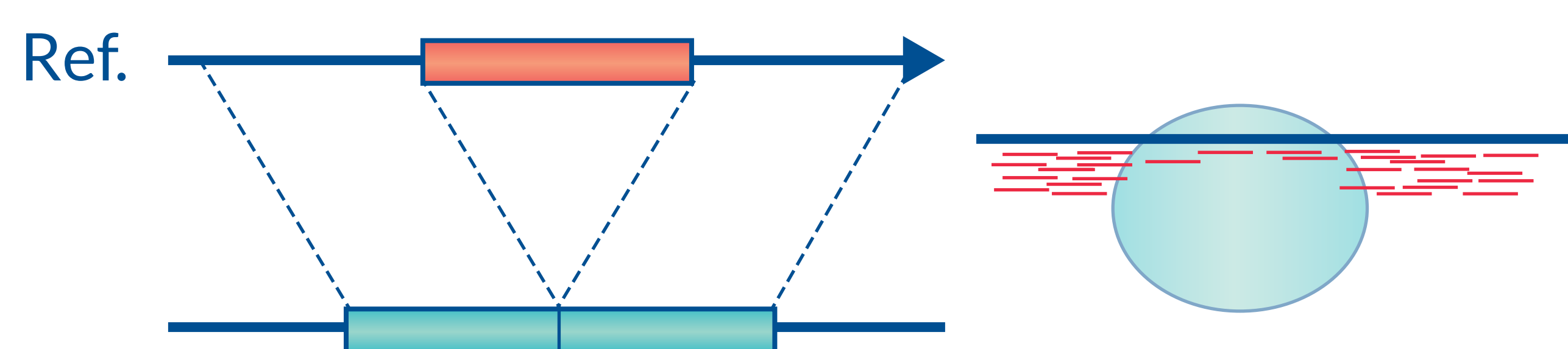
## INTRODUÇÃO E MOTIVAÇÃO

O diagnóstico molecular de doenças genéticas é, atualmente, ferramenta indispensável para o seguimento clínico adequado dos pacientes e para o aconselhamento genético familiar. Variações no número de cópias (CNVs) são segmentos do DNA que foram deletados ou duplicados em tamanho maior que 50 pares de bases, podendo representar até 10% dos casos de câncer hereditário e 14% dos casos de doenças genéticas pediátricas [1, 2, 3].

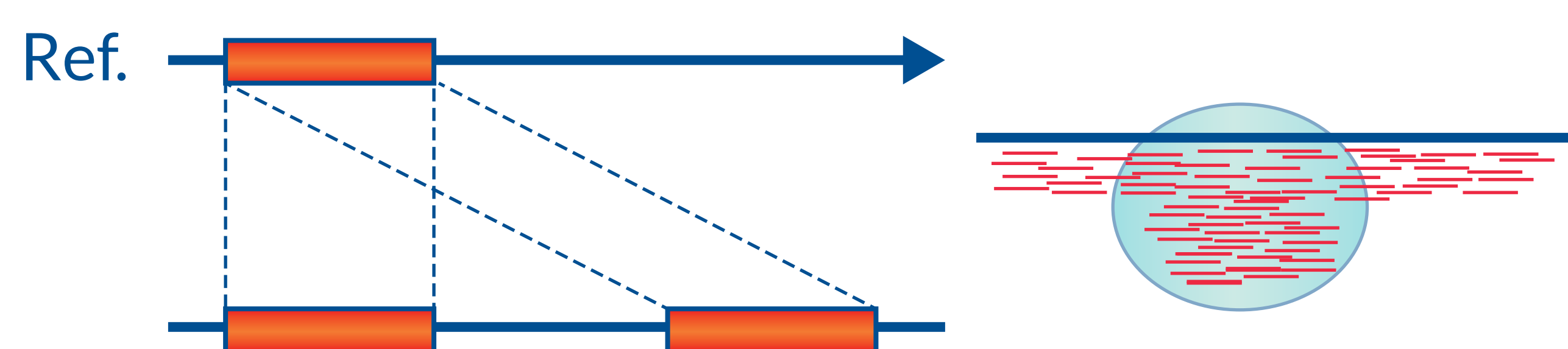
Com o custo e a complexidade na análise dos dados gerados pelo sequenciamento de genoma completo (WGS), o sequenciamento de exoma (WES) e o de amplicons têm se tornado as principais metodologias com propósito diagnóstico usando sequenciamento de nova geração (NGS). Essas técnicas têm demonstrado alta capacidade na identificação de variantes de nucleotídeo únicos (SNVs) e pequenas inserções/deleções (INDELS) raras, porém a identificação de grandes deleções e duplicações são menos confiantes nessas técnicas quando comparadas a utilização do WGS, da hibridização genômica comparativa baseada em microarranjos (aCGH) e da amplificação de múltiplas sondas dependentes de ligação (MLPA) [4].

Para que painéis genéticos e o sequenciamento de exoma substituam essas outras metodologias é necessário um software ou a combinação de softwares que detectam CNVs por exons com alta sensibilidade, especificidade e uma taxa de falso-positivos aceitável. Com isso, iniciamos o desenvolvimento de pipelines para a detecção de CNVs em painéis e exomas possibilitando o aumento de casos clínicos elucidados.

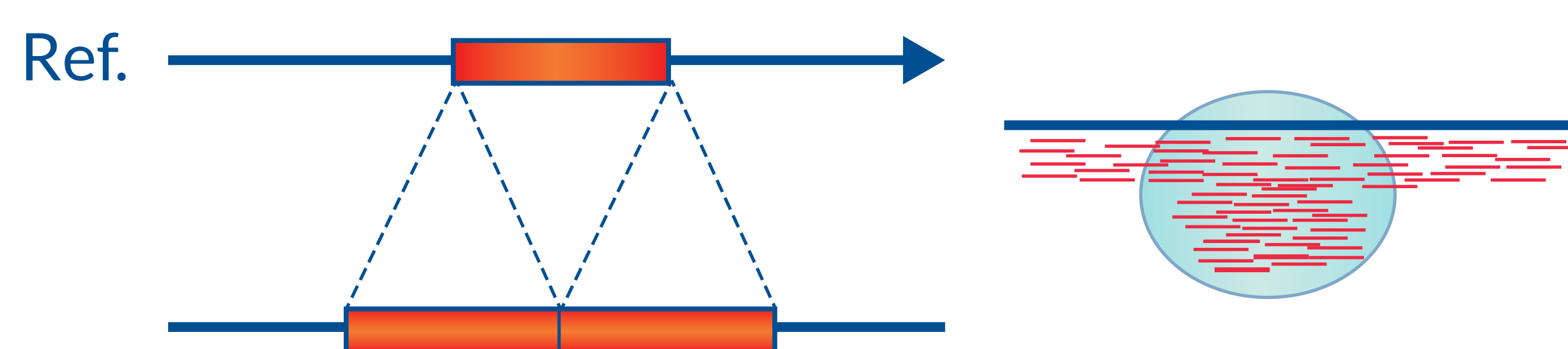
### DELEÇÃO



### DUPLICAÇÃO INTERCALADA



### DUPLICAÇÃO EM TANDEM



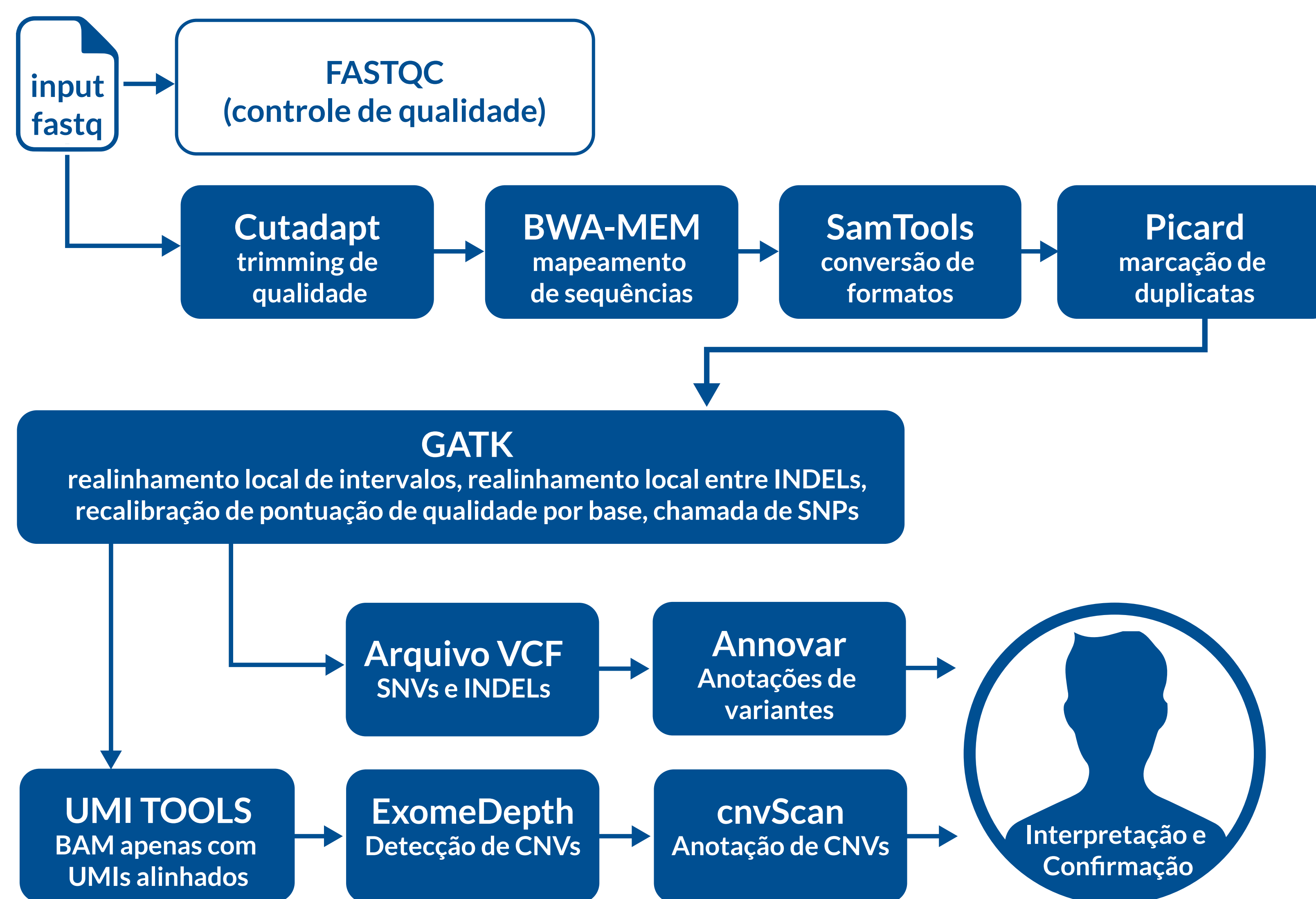
**Figura 1:** Classes de CNVs possíveis de serem identificadas usando algoritmos de leitura de cobertura disponíveis para sequenciamento de amplicons.

## REFERÊNCIAS

1. Alkan, Can, Bradley P. Coe, and Evan E. Eichler. "Genome structural variation discovery and genotyping." *Nature Reviews Genetics* 12.5 (2011): 363.
2. Shlien, Adam, and David Malkin. "Copy number variations and cancer." *Genome medicine* 1.6 (2009): 62.
3. Cooper, Gregory M., et al. "A copy number variation morbidity map of developmental delay." *Nature genetics* 43.9 (2011): 838.
4. Ellingford, Jamie M., et al. "Validation of copy number variation analysis for next-generation sequencing diagnostics." *European Journal of Human Genetics* 25.6 (2017): 719.
5. Carvalho, George; et al. "Improving variant accuracy with copy number variant pipeline for target sequencing" - <https://bit.ly/2jRNKC5>
6. Smith, Tom, Andreas Heger, and Ian Sudbery. "UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy." *Genome research* 27.3 (2017): 491-499.

## MÉTODOS

As amostras foram sequenciadas na Genomika Diagnósticos utilizando o kit Qiagen V3, que usa em seu protocolo barcodes moleculares. Esta etapa possibilita uma melhor acurácia na descoberta de SNVs e mantém uma homogeneidade na quantidade de leituras observadas. Através de um benchmark prévio entre o ExomeDepth e o CoNVaDING foi possível selecionar o primeiro software a ser utilizado em nossas rotinas. O ExomeDepth obteve 100% de acurácia, precisão, sensibilidade e especificidade em nosso estudo prévio. Porém, com o uso em rotina foi possível perceber um aumento gradual na taxa de falso-positivos [5]. Todas as amostras utilizadas no benchmark anterior somadas as amostras confirmadas como falso-positivas por MLPA (n=29), foram reprocessadas após a introdução do software UMItools que permite extrair apenas os barcodes moleculares únicos para a análise de cobertura [6].



**Figure 2:** Processos internalizados para a identificação de CNVs usando painéis genéticos.

## BENCHMARK

Pipeline	VP	VN	FP	FN	Acurácia	Precisão	Sensibilidade	Especificidade
UMIs duplicados	12	15	2	0	93,10%	85,71%	100%	88,23%
Apenas UMIs únicos	12	17	0	0	100%	100%	100%	100%

**Tabela 1:** Benchmark entre antes e depois da implantação do UMItools. Foram usados cálculos estatísticos derivados da matriz de confusão. Contagem de verdadeiro-positivos (VP), verdadeiro-negativos (VN), falso-positivos (FP) e falso-negativos (FN) foi baseada na quantidade amostras com confirmações por MLPA (n=29).

## DISCUSSÃO E CONCLUSÃO

Com a introdução do processo de extração de barcodes moleculares foi possível diminuir a taxa de falso-positivos para zero dentro do conjunto de amostras com confirmações de MLPA disponíveis no laboratório. Nossos resultados demonstram a importância da utilização de ferramentas para a detecção de CNVs na rotina clínica, limitando o uso de metodologias padrão-ouro como o MLPA apenas para as amostras com variantes encontradas com pipeline de bioinformática para detecção de CNVs. Todavia, é necessário realizar mais validações com um número maior de amostras testadas antes de implementar totalmente essa metodologia. Adicionalmente, como indicado nas boas práticas para bioinformática clínica, pretendemos testar outras ferramentas para que se tenha mais de um software para detecção de CNVs com respaldo clínico.