

作品名稱
適用於次世代定序識別基因變體之系統單晶片

A Fully Integrated Genetic Variant Discovery SoC for Next-Generation Sequencing

隊伍名稱
魯蛇肥肥就愛吃 Fat Loser Consumes a Lot

隊長
吳易忠 臺灣大學電子工程學研究所

隊員
陳彥龍 臺灣大學電子工程學研究所
楊仲萱 臺灣大學電子工程學研究所



作品摘要

此作品為全世界第一個支援完整次世代定序識別基因變體之系統。此系統搭配本團隊所設計與下線之系統單晶片與客製化電路板、周邊電路，達到全世界最快運算速度。相較於高階顯卡有 66 倍之加速幅度。本作品之系統單晶片可支援四種運算，資料前處理、短片段回貼、半倍體搜尋與變體識別。

透過本團隊所開發之 sBWT 演算法，搭配現有之基因組分析套件，此作品之精準度可與軟體平台一致。於所設計之系統單晶片中，本作品透過多種硬體優化技巧，如高度平行、硬體共享、複雜化簡以達到高效能與低功耗之特性。此晶片亦設計了兩個主要運算單元，分別為多工排序引擎與動態規劃處理引擎來支援整個基因定序資料分析之複雜運算。

本系統亦整合一顆 Synopsys ARC 處理器，可用於檔案傳輸介面、記憶體資料與 IP 控制等等，以增加系統彈性。此晶片透過臺積電 28nm 製程下線，可操作在最高 400MHz，功耗為 0.975W，並可在 37 分鐘內完成完整基因資料分析。與高階顯卡相比，不僅速度有顯著提升，於能量效率與面積效率上亦有數個量級以上的增益。

本系統經由 FDA 之標準測資完成驗證，可達到 99.6% 之精確度，所設計之客製化 GUI 亦可達到即時判讀之需求。

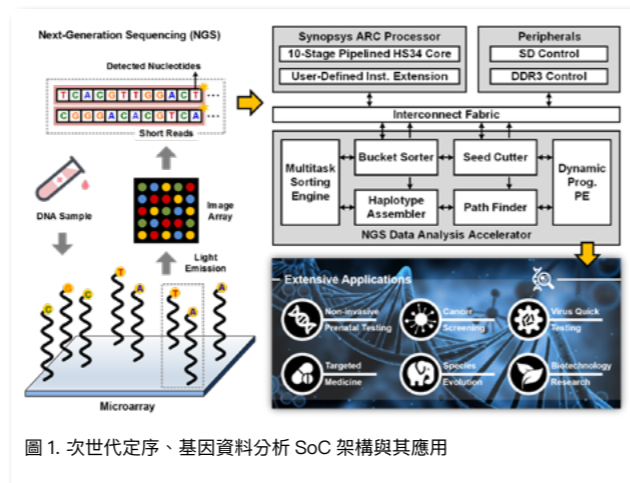
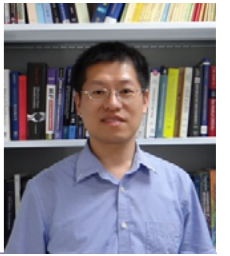


圖 1. 次世代定序、基因資料分析 SoC 架構與其應用

楊家驥 臺灣大學電機工程學系

- 美國加州大學洛杉磯分校電機博士，現為臺灣大學電機工程學系教授。實驗室致力於開發低功耗之客製化晶片以提升資料處理速度與能量效率。
- 研究領域：生醫訊號處理器、基頻通訊積體電路、與機器學習處理晶片設計



洪瑞鴻 交通大學資訊工程學系

- 美國波士頓大學博士，現為交通大學資訊工程學系副教授，實驗室致力於開發更高效的演算法與計算模型來了解基因體的複雜問題。
- 研究領域：次世代定序演算法、生物資訊分析、複雜系統與機器學習



Abstract

This work presents the first complete system with dedicated system-on-chip (SoC) that supports full data analysis workflow for genetic variant discovery for next-generation sequencing (NGS). The SoC implements four major steps: preprocessing, short-read mapping, haplotype calling, and variant calling.

By adopting the sBWT-based short read mapping and Genome Analysis ToolKit (GATK) Haplotype Caller (HC)-based variant calling algorithms, this work achieves a comparable precision as the software solutions. The dedicated hardware architecture achieves high parallelism with low hardware complexity. In this SoC, the multitask sorting engine (MTSE) and the dynamic programming processing engine (DPPE) are proposed for essential computing tasks for NGS data analysis.

An ARC processor is integrated to facilitate IO interfaces, including a DDR3 PHY. Fabricated in a 28nm technology, the chip area is 12 mm². It dissipates 975 mW at a clock frequency of 400 MHz from a 0.9-V supply. The SoC is able to complete variant discovery for the whole human genome within 37 minutes. Compared to an optimized GPU solution, this work is 66× faster, and also achieves an 1.59×10⁴ times higher energy efficiency and 3,086× higher area efficiency.

Verified with the FDA standardized benchmark, the SoC achieves a precision of 99.6%. The prototype system demonstrates the analysis procedure in real-time for on-site DNA sequencing

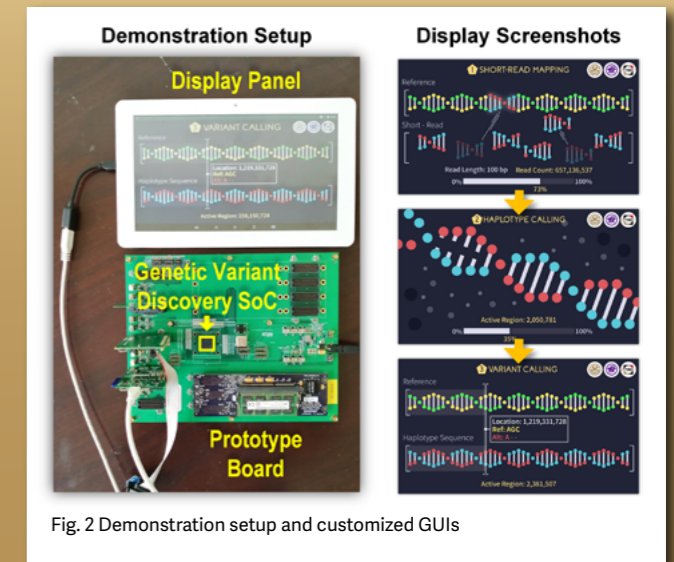


Fig. 2 Demonstration setup and customized GUIs