

kontron + intel®

Germline Variant Calling on the Kontron CG2600 with Intel® Xeon® 6



A whole human genome, deep-learning variant calling, in under 26 minutes on a quiet air-cooled bench server.

No GPU, no accelerator card, no facilities project.

25m 47s
HG002 35x
whole genome

3.79x
16 to 96 cores,
identical output

68 dB
at sustained
100% load

Summary

Genomics teams increasingly want deep-learning variant calling close to the sequencer rather than queued behind shared infrastructure. This brief describes a bench-side configuration from Kontron and Intel running DeepVariant 1.10 germline variant calling on Intel® Xeon® 6 processors. A pre-aligned HG002 35x whole genome completes in 25 minutes 47 seconds. The deep-learning inference runs on Intel® AVX-512 — the vector units already present in every Xeon core. Nothing is added to the bill of materials, and nothing is added to the facility.

► Acceleration without procurement

DeepVariant 1.10 runs an fp32 TensorFlow model, and AVX-512 executes those convolutions directly on the CPU. No add-in cards, no accelerator lead time, no CUDA-specific stack to support.

► Core count and sharding

Variant calling scales with core count — 3.79x going from 16 to 96 physical cores. Both configurations produced byte-identical output: the same 7,709,239 variants, 6,443,505 SNPs and 1,265,734 indels. Core count changes the wall clock and nothing else.

► Capacity is a purchase, not a re-engineering project

More cores means more throughput with no code changes and no pipeline rewrite.

► Runs fully offline

Pipeline containers run with networking disabled entirely. The genome never leaves the box — often a hard regulatory requirement rather than a preference.

► Siting and acoustics

Air-cooled and bench-deployable — no plumbing, no isolated server room, no facilities project. Deploy it next to the sequencer.

➤ Available configuration

Feature	Configuration
Platform	Kontron CG2600, dual-socket, air-cooled rack server — lab-bench deployable
CPU	2x Intel® Xeon® 6740P — 48 cores per socket, 96 cores / 192 threads total
Acceleration	Intel® AVX-512 vector units, built into every core — no add-in card
DRAM	1 TB DDR5
Storage	U.2 NVMe data disks; 960 GB M.2 NVMe boot
Operating system	Ubuntu 24.04 LTS with Docker
Workload	google/deepvariant:1.10.0, run_deepvariant BAM-in path, 192 shards
Data	HG002 (GIAB), NovaSeq PCR-free 35x, GRCh38 no-ALT analysis set

➤ Benchmark — HG002 35x whole genome, BAM to VCF

Configuration	Cores used	Wall clock	Variants called
CG2600 — all 192 threads	96C / 192T	25m 47s	7,709,239
CG2600 — 96 physical cores	96 cores	32m 59s	7,709,239
CG2600 — 16 physical cores	16 cores	2h 04m 51s	7,709,239

Quickstart

Run a complete whole-genome variant-calling workflow on the Kontron CG2600 with Intel® technology and Google DeepVariant.

HG002 whole genome | BAM to VCF | 25m 47s | 7.7M+ variants called

Learn more

Contact your **Kontron** or **Intel** account team for a tailored configuration.

Reference data:

GIAB HG002 and the GRCh38 no-ALT analysis set, both public. DeepVariant is open source at github.com/google/deepvariant.

ASHG 2026

Montréal • October 20–24 • Booth #1204

Book a live demo

