

Cancer Genomics in Galaxy

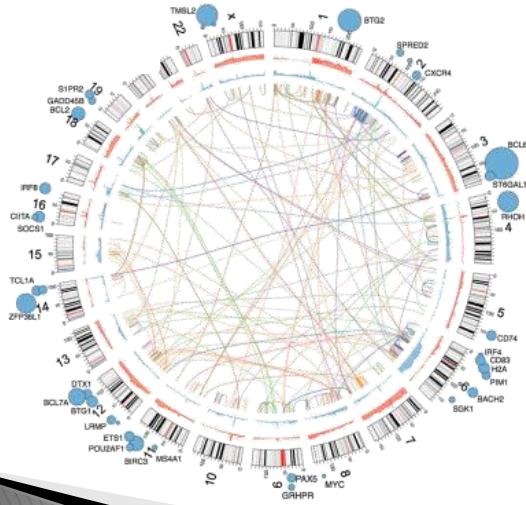
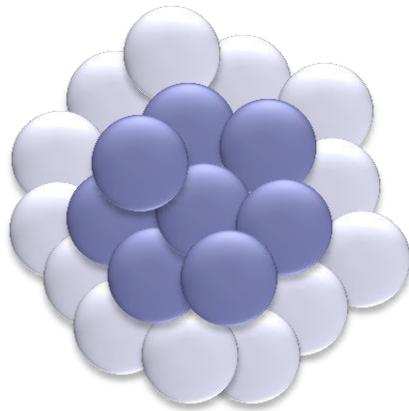
Marco Albuquerque

Morin Laboratory

Simon Fraser University

British Columbia, Canada

Cancer Genomics



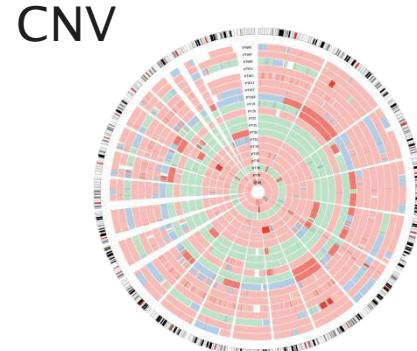
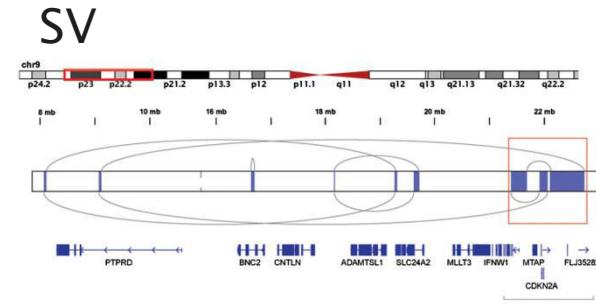
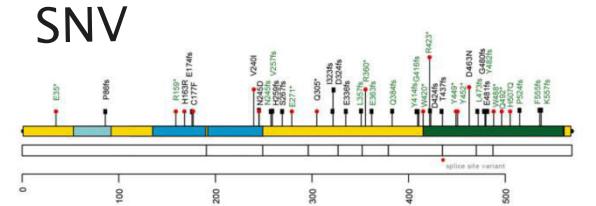
A Bottleneck on Data Analysis



Avoid cryptic command line options



Ensure easy installation



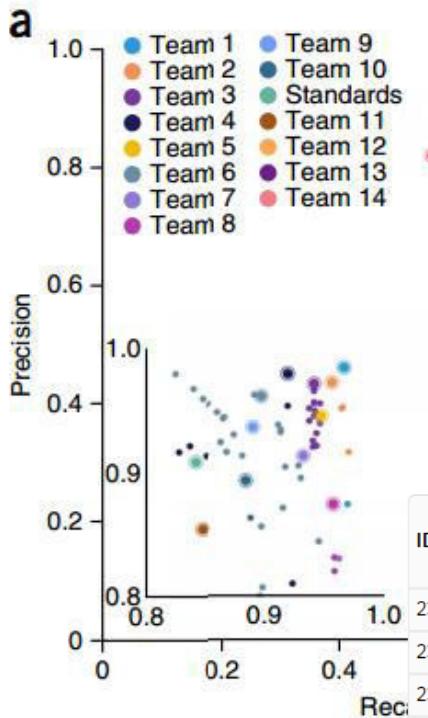
Available
anywhere



Challenges

- ▶ Tool Benchmarking
- ▶ Optimal Parameterization
- ▶ Economical
- ▶ Parallelization

Informed Tool Parameterization

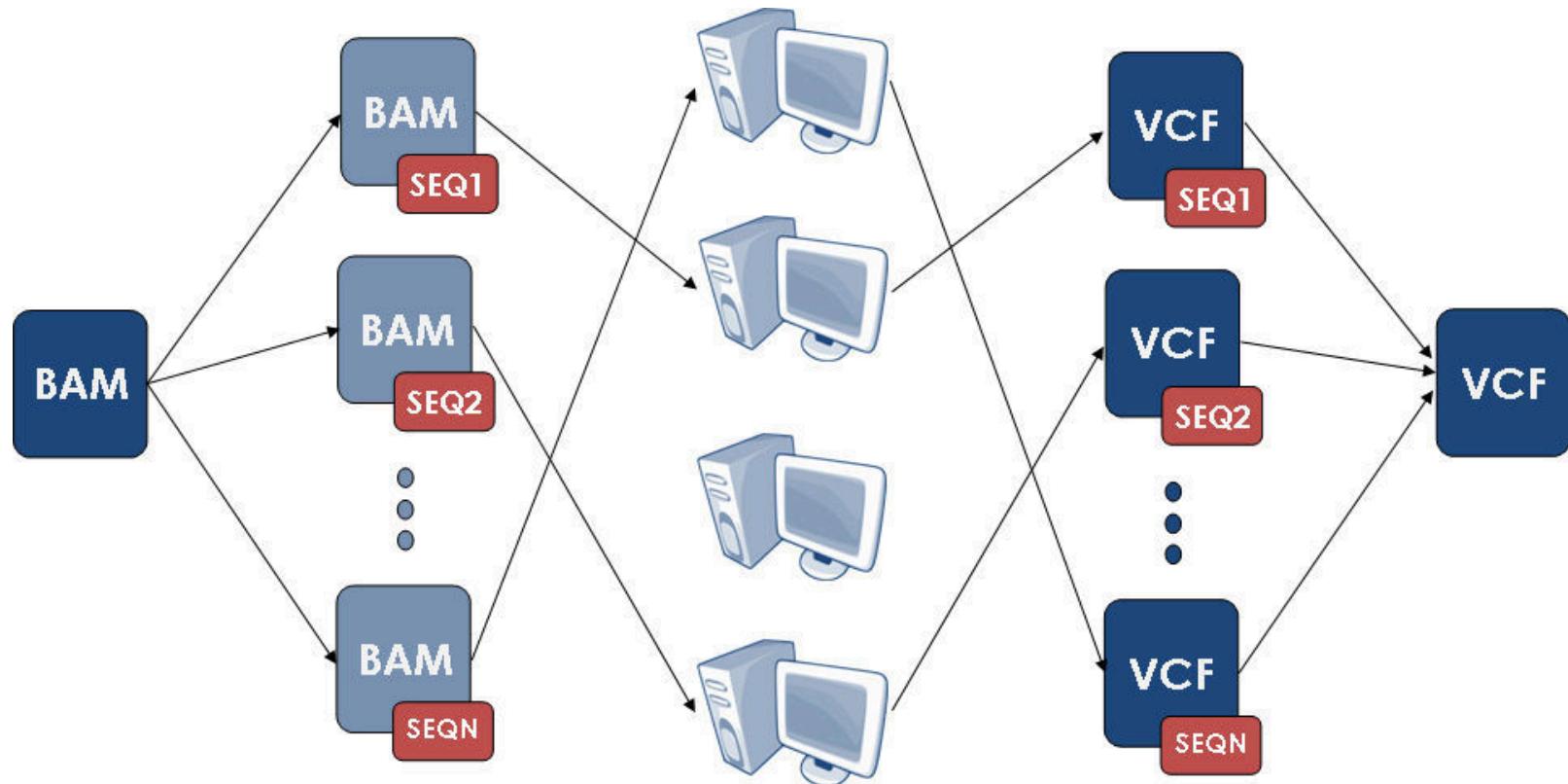


Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection

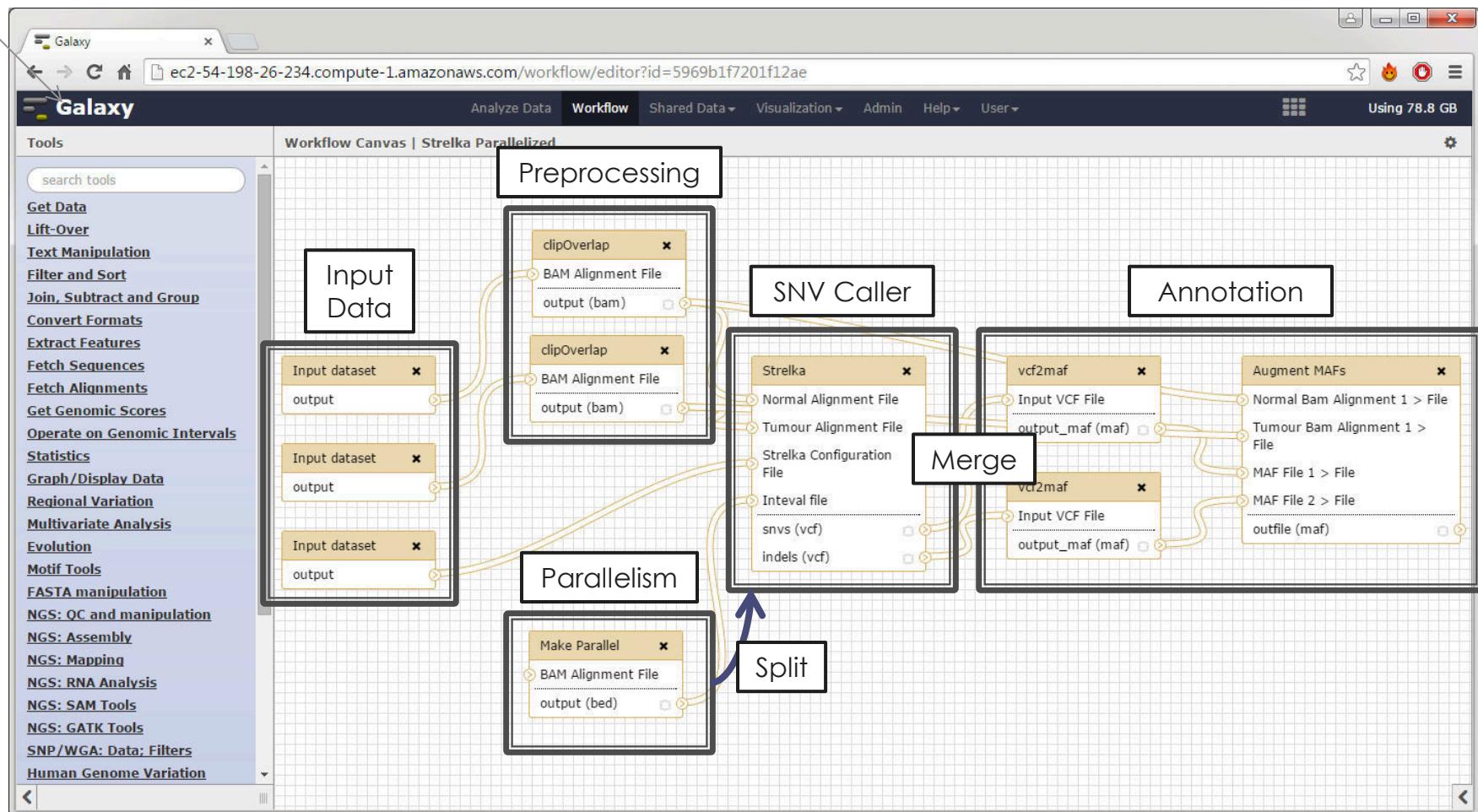
Adam D Ewing^{1,2,11}, Kathleen E Houlahan^{3,11}, Yin Hu^{4,11}, Kyle Ellrott¹, Cristian Caloian³, Takafumi N Yamaguchi³, J Christopher Bare⁴, Christine P'ng³, Daryl Waggett³, Veronica Y Sabelnykova³, ICGC-TCGA DREAM Somatic Mutation Calling Challenge participants⁵, Michael R Kellen⁴, Thea C Norman⁴, David Haussler¹, Stephen H Friend⁴, Gustavo Stolovitzky⁶, Adam A Margolin^{4,7,8,12}, Joshua M Stuart^{1,12} & Paul C Boutros^{3,9,10,12}

ID	Submission Name	Team	Status	SMC Challenge Eligible	Number of Calls (Censor:Chr1)	Number of Calls (All)	Sensitivity: (Censor:Chr1)	Sensitivity: (All)	Precision: (Censor:Chr1)
2322158	BreakDancer.set1.test	SMC_Admins	SCORED	YES	1122450	1215967	0.23121	0.22102	0.00007
2342351	default destruct run	Shah Lab	SCORED	YES	5673	6216	0.82370	0.82480	0.54680
2342937	Delly default-params	DellyTeam	SCORED	YES	296	317	0.81792	0.81402	0.95608
2343008	Only Somatic destruct run	Shah Lab	SCORED	YES	242	270	0.23988	0.25067	0.83884
2343125	destruct somatic sscl=0	Shah Lab	SCORED	YES	2714	2966	0.82370	0.82480	0.59912
2343130	destruct no spl2	Shah Lab	SCORED	YES	5673	6215	0.82370	0.82480	0.54680
2343132	destruct somatic no spl2	Shah Lab	SCORED	YES	2714	2965	0.82370	0.82480	0.59912
2343136	destruct somatic no spl2 score > 0	Shah Lab	SCORED	YES	2692	2943	0.82370	0.82480	0.59993
2343138	destruct somatic no spl2 score > 5	Shah Lab	SCORED	YES	1860	2031	0.80058	0.80323	0.64570

Parallelization in Galaxy



Somatic SNV Workflow



Current Tools

Strelka
Mutect
SNV
MutationSeq

TITAN
sequenza
CNV
ExomeCNV

Lumpy
Pindel
SV DELLY
FACTERA

Circos
Visualization
Plot Protein

Metric
Calculator
Purity Extras
Estimator
Annotators

Acknowledgements

The Boutros Lab

Anna Lee
Vincent Huang
Takafumi Yamaguchi
Dr. Paul Boutros

The Shah Lab

Dr. Cydney Nielsen
Diljot Grewal
Jafar Taghiyar Renani
Dr. Sohrab Shah

The Morin Lab

Arezoo Mohajeri
Bruno Grande
Dan Fornika
Elie Ritch
Jasleen Grewal
Jordan Davidson
Kevin Bushell
Kristina Aluzaitė
Dr. Miguel Alcaide Torres
Sarah Arthur
Dr. Sepideh Alamouti
Stephen Yu
Dr. Ryan Morin



GenomeCanada

