Assessing tumor heterogeneity and tracking clonal evolution using whole genome or exome sequencing

Chris Miller, PhD
Tumors are heterogeneous

genetically diverse populations of cells

evolution occurs at the cellular level
Clonal evolution in relapsed AML

Nature 481, 506–510 doi:10.1038/nature10738
Challenges for detecting minor subclones

- Genomes are sequenced with low coverage
  - 30x not enough

- Algorithms aren’t designed to detect low-frequency events
Somatic Sniper power simulations

• 90x coverage

• Power to detect event at 20% VAF: 85%

• Power to detect event at 10% VAF: 10%

BASSOVAC

- **BAyesian Scoring of Somatic VAriant read Counts**

- Incorporates purity, ploidy, base quality, allele frequency, and overall mutation rate.

- Bayesian framework for inversion to obtain, probabilities of heterozygous and homozygous somatic events, given the data.
Simulation

90X tumor 30X normal, Q20 simulation data, BMR 1e-6
with mapping and false-positive filtering

van Rijsbergen’s F1 score

F1 = \frac{2 \times \text{prec.} \times \text{recall}}{\text{prec.} + \text{recall}}

nominal DWGSIM variant allele frequency
Real-world testing

- Primary breast tumor
- Matched normal
- 3 different metastases:
  - Spinal
  - Liver
  - Adrenal
Real-world testing

• All whole-genome sequenced to 30x
• Mutation calls made with Somatic Sniper and Varscan
• Capture validation performed for all variants
• Deep readcounts obtained from validation sequencing for all variants in all samples
BRCA Tumor vs Spinal Metastasis
BRCA Tumor vs Spinal Metastasis
BRCA Tumor vs Spinal Metastasis
BRCA Tumor vs Spinal Metastasis
BRCA Tumor vs Spinal Metastasis
BRCA Tumor vs Spinal Metastasis

Present in the tumor at low frequency
BASSOVAC sensitivity - BRCA met SNVs in tumor

The graph shows the number of true events detected against tumor variant allele frequency for three different methods: Bassovac (1990 TP), Sniper (1703 TP), and Strelka (956 TP). The graph indicates a peak in detection at around 20% allele frequency for all three methods, with Bassovac showing the highest detection rate.
BASSOVAC

- Over 50% of the variants present in the metastases are present at a detectable level in the tumor

- We can use BASSOVAC to detect true variants at very low frequencies (< 2%)
Clonal inference

• Given information about a tumor, how many clones are present?
  • Which variants are present in different subclones?

• Requires integrative approach
  • Variant allele frequencies
  • Copy number calls
  • Purity and Ploidy information
Clonality Plot
Clonality Plot

Kernel Density

SNVs in CN2 regions

SNVs in CN3 regions
Infer clones in an automated, unbiased manner
Biclonal sample

TCGA-B5-A0JV-01A-11D-A10B-09

Tumor Variant Allele Frequency

Kernel Density

Tumor Coverage

2 Copies

18.4

18.3

43.4
Triclonal sample

TCGA-AX-A063-01A-11W-A027-09

Tumor Variant Allele Frequency

Kernel Density

Tumor Coverage

11.9 22.9 38.3
Non-intuitive sample

TCGA-BG-A0MQ-01A-11D-A10B-09

Tumor Variant Allele Frequency

Kernel Density

Tumor Coverage

12.2 21.9 42.6
Multi-clonal Sample

TCGA-D1-A15X-01A-11D-A122-09

Tumor Variant Allele Frequency

Kernel Density

2 Copies

Tumor Coverage

13.3, 8.7, 2932.57, 12.2
Clonal inference

- Most tumors have a founding clone and one or more subclones (LAML, BRCA, UCEC)
- Lower bound on number of clones
Conclusions

• We can detect somatic mutations at very low frequencies using BASSOVAC

• We have developed robust automatic methods for inferring details about the subclonal architecture of a tumor

• Goal: characterizing minor subclones at diagnosis, rather than discovering their presence at relapse
Acknowledgements

- Mike Wendl
- Nathan Dees
- Dave Larson
- Travis Abbott
- Beifang Niu
- Brian White
- Will Schierding
- Josh McMichael
- Charles Lu
- Krishna Kanchi
- Tim Ley
- Michael Tomasson
- Ramaswamy Govindan
- Matthew Ellis
- Chuck Perou
- Elaine Mardis
- Rick Wilson
- Li Ding
- NHGRI
- NCI

The Cancer Genome Atlas

Understanding genomics to improve cancer care