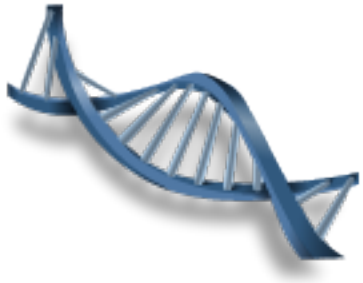
Extraction of Human DNA Replication Timing Patterns from Discrete Microarray Data



Anindya Dutta, Neerja Karnani,
Ankit Malhotra, Gabriel Robins,
and **Christopher M. Taylor**

UNIVERSITY of VIRGINIA

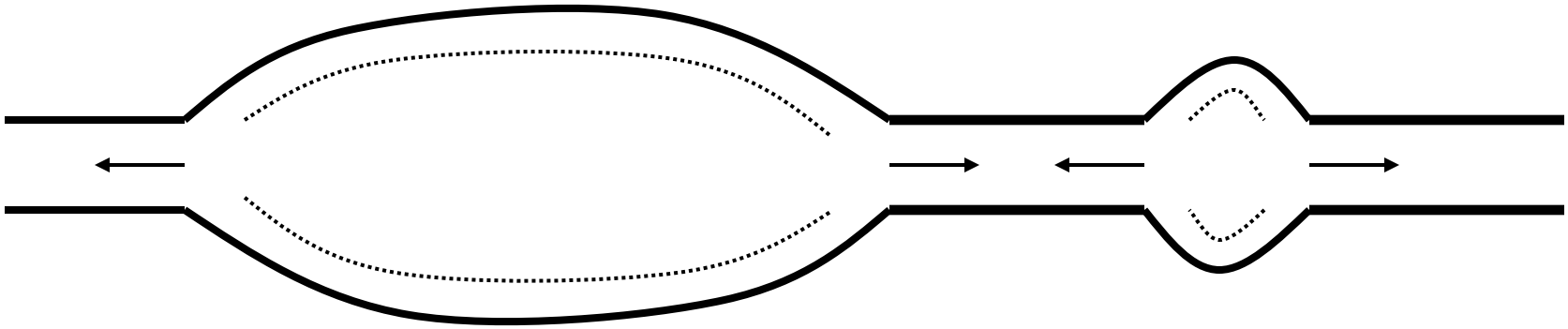
UNIVERSITY of NEW ORLEANS

Outline

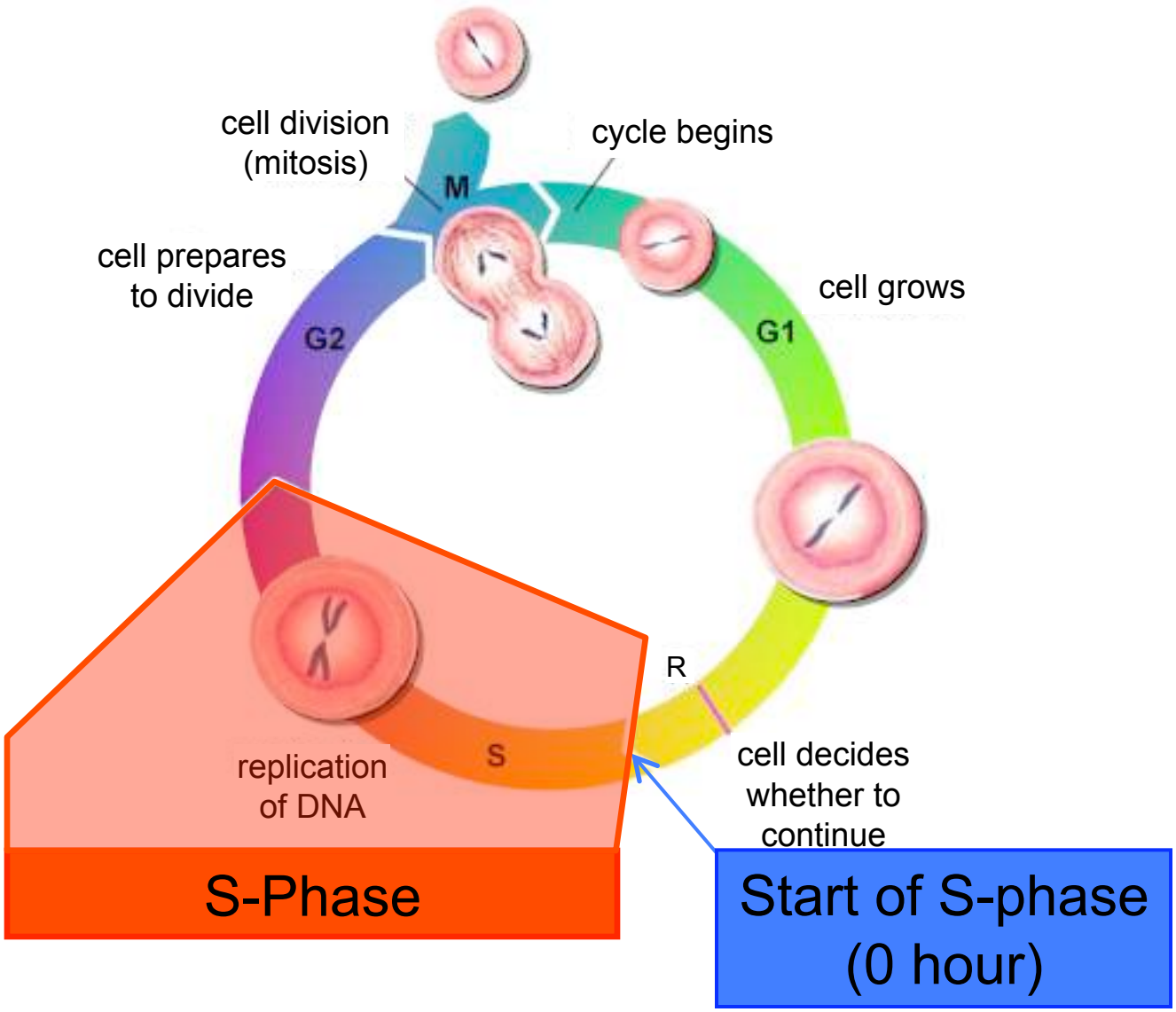
- Introduction
- Data Collection
- Temporal Specificity and Allelic Variation
- Time of Replication of 50% (TR50) of a Locus
- Smoothing
- Segregation
- Results
- Summary
- Acknowledgements and Data Availability
- Collaboration Avenues

DNA Replication

- A crucial step in the cell cycle
  - Passing on genetic information
- Replication initiates at origins
  - Origins fire at different times during S-phase
- Goal: Profile DNA replication timing

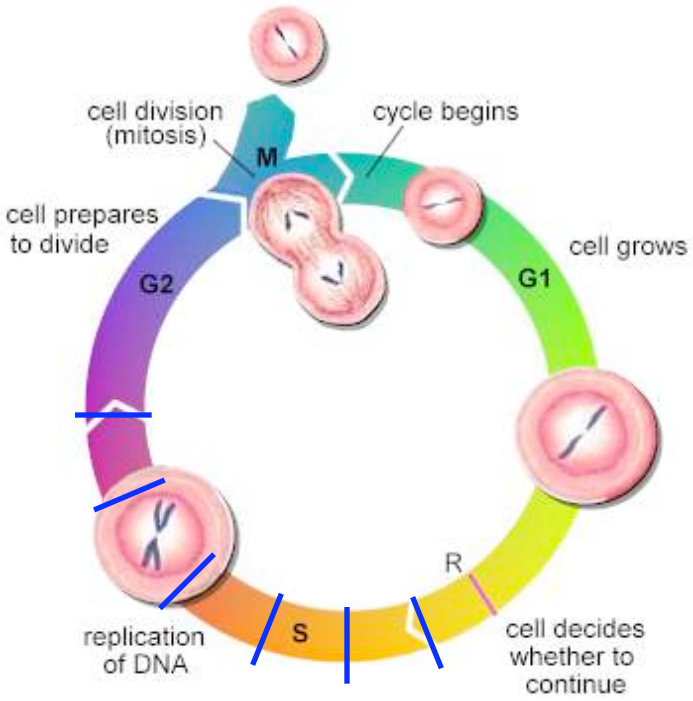


The Cell Cycle



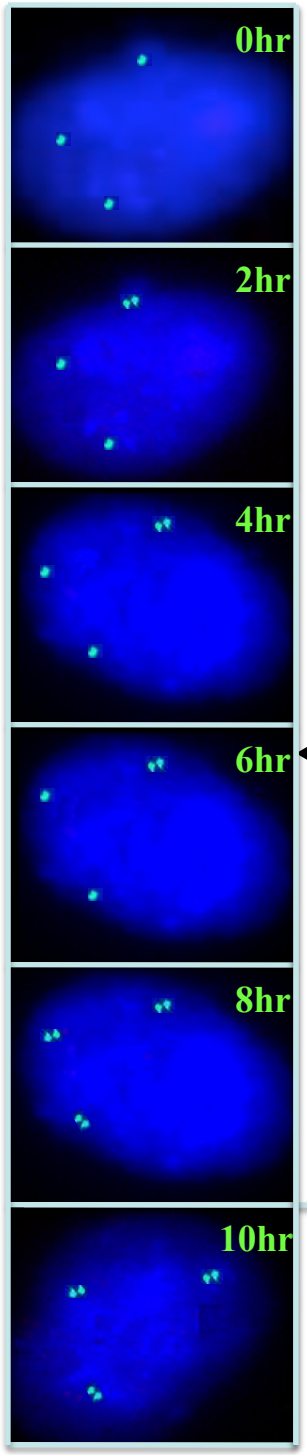
Profiling DNA Replication Timing

- Ideal: f(chr, bp) = rtime
- Isolate DNA replicated in discrete parts of S-phase
  - One cell is not enough
  - Synchronize S-phase entry
    - Apply drugs
    - Release together
      - Synchronization error
  - Label in intervals of duration *L*
- Allelic Variation
  - mf(chr, bp) = {rtime1, rtime2, …}



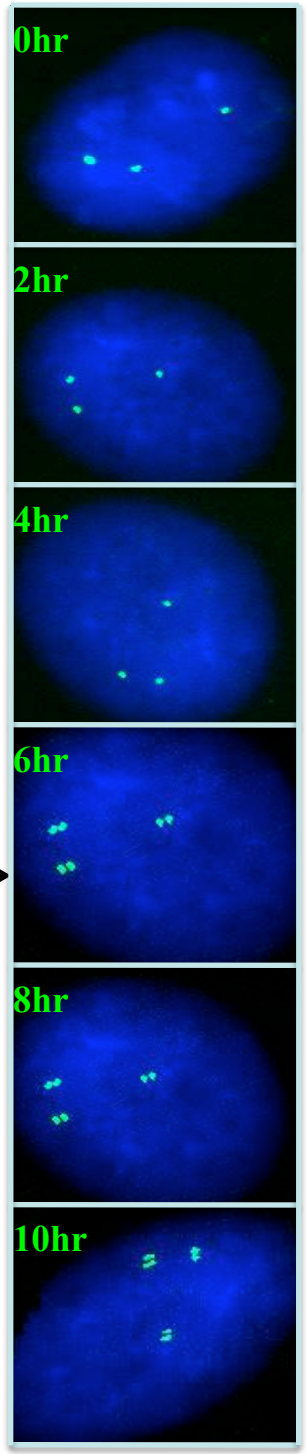
Allelic Variation

- Fluorescent in-situ Hybridization (FISH)
  - Replication timing at a given site
  - HeLa cells show 3 alleles at locus

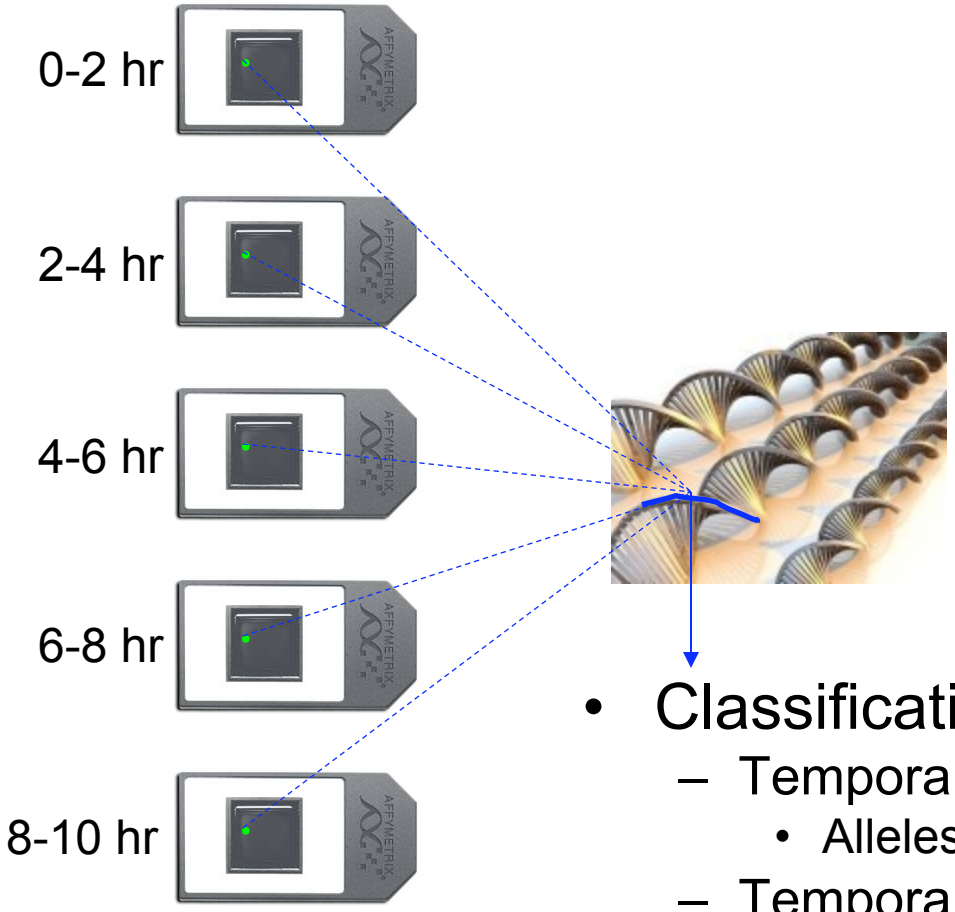


Temporally non-specific replication (TNS)

Temporally specific replication (TS)



Initial Processing

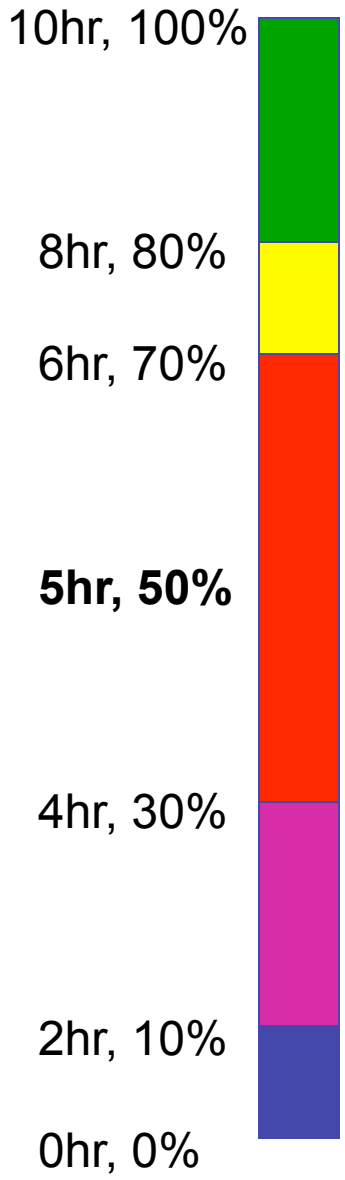


// Is there evidence that all alleles are replicating together?
If (max sum of two adjacent time periods > (1 - 1/N) * total*)*
then {probe is temporally specific}
// Is at least one allele replicating apart from the majority?
Else If (max sum of two adjacent time points not including the maximum time point ≥ 1/N * total)
then {probe is temporally non-specific}
// Isolated signal is not strong enough to be an allele.
Else
{probe is temporally specific}

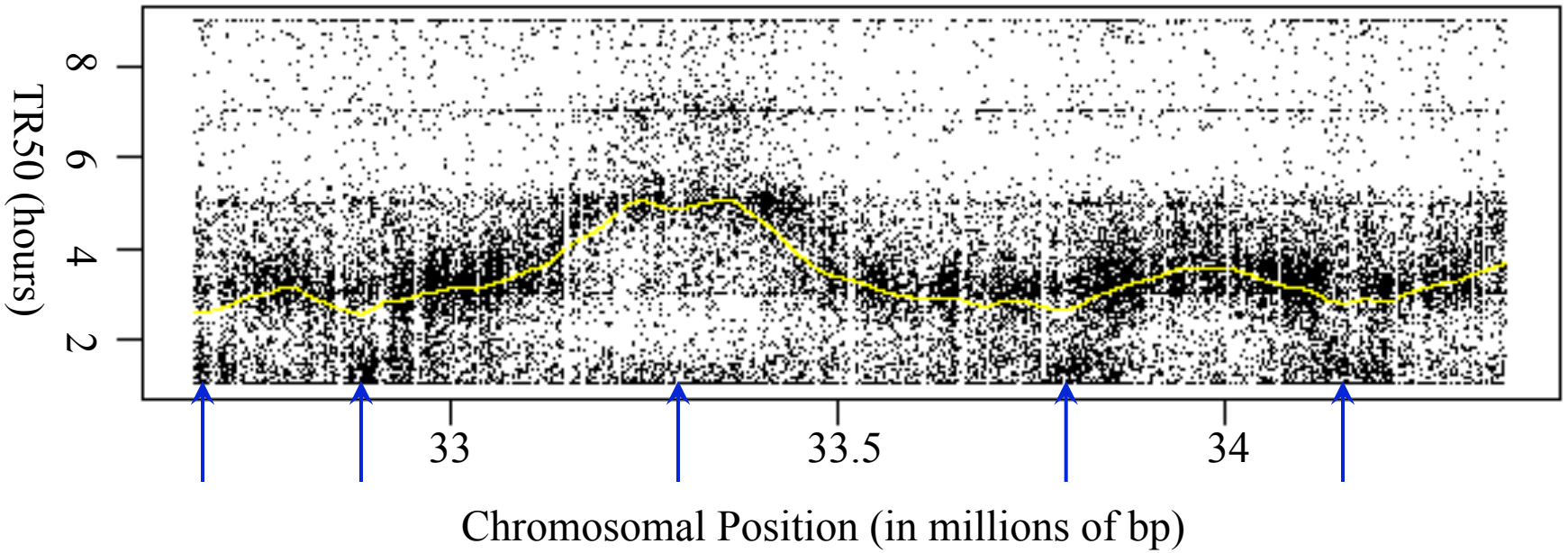
- Classification Algorithm
  - Temporally Non-Specific Replication (TNSR)
    - Alleles replicate separately
  - Temporally Specific Replication (TSR)
    - Time of Replication of 50% (TR50)

Time of Replication of 50% (TR50)

- Computed for each probe on array
- Linearly interpolate the time that 50% of a probe's signal occurred
- Discard probes with no signal
- Example →
  - TR50 value occurs at **5hr**

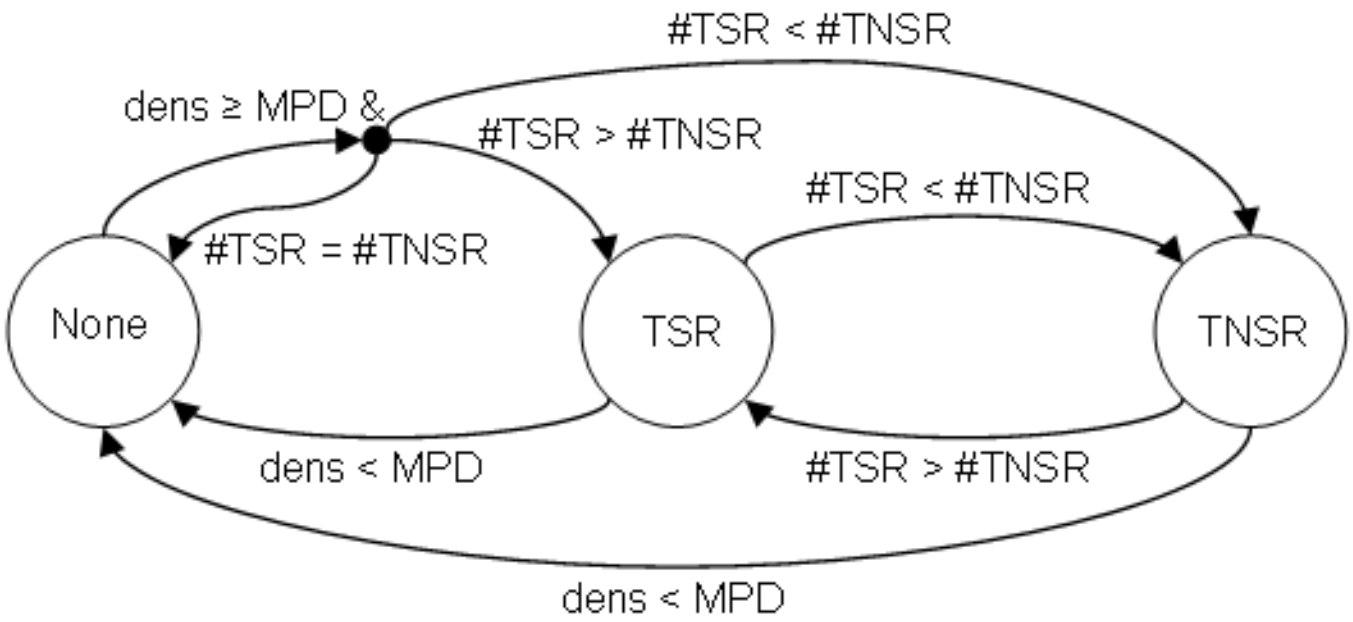


Plotting TR50



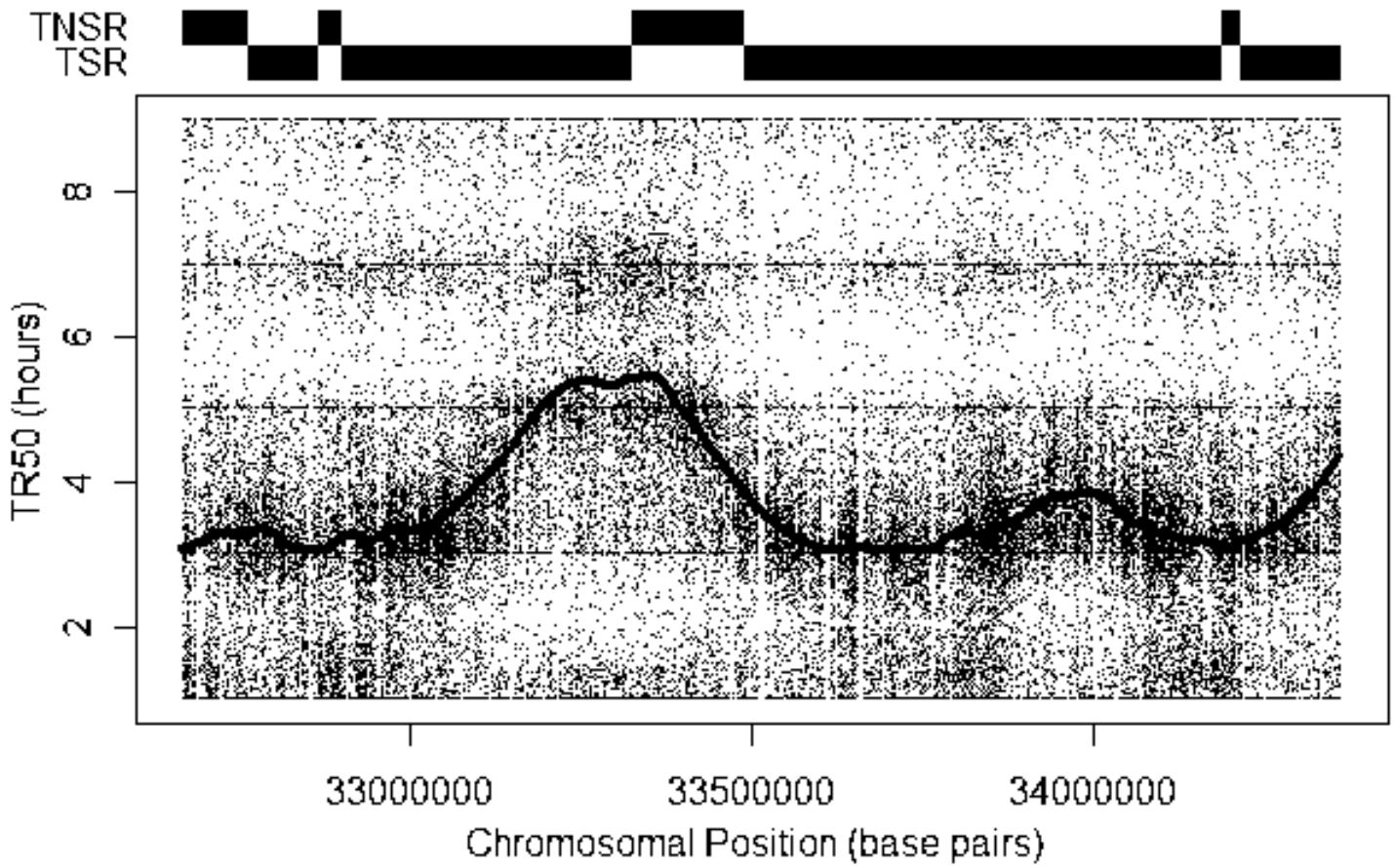
- Smoothed TR50 curve recovers replication pattern
- Local minima → Possible locations of replication origin

Segregation Algorithm



- Sliding window passes over probes to generate intervals
  - Density (dens) of probes in window ≥ MPD ⇒ Generate Interval
  - Ratio of TSR to TNSR probes determines temporal specificity
- Generates two disjoint sets of intervals (TSR and TNSR)

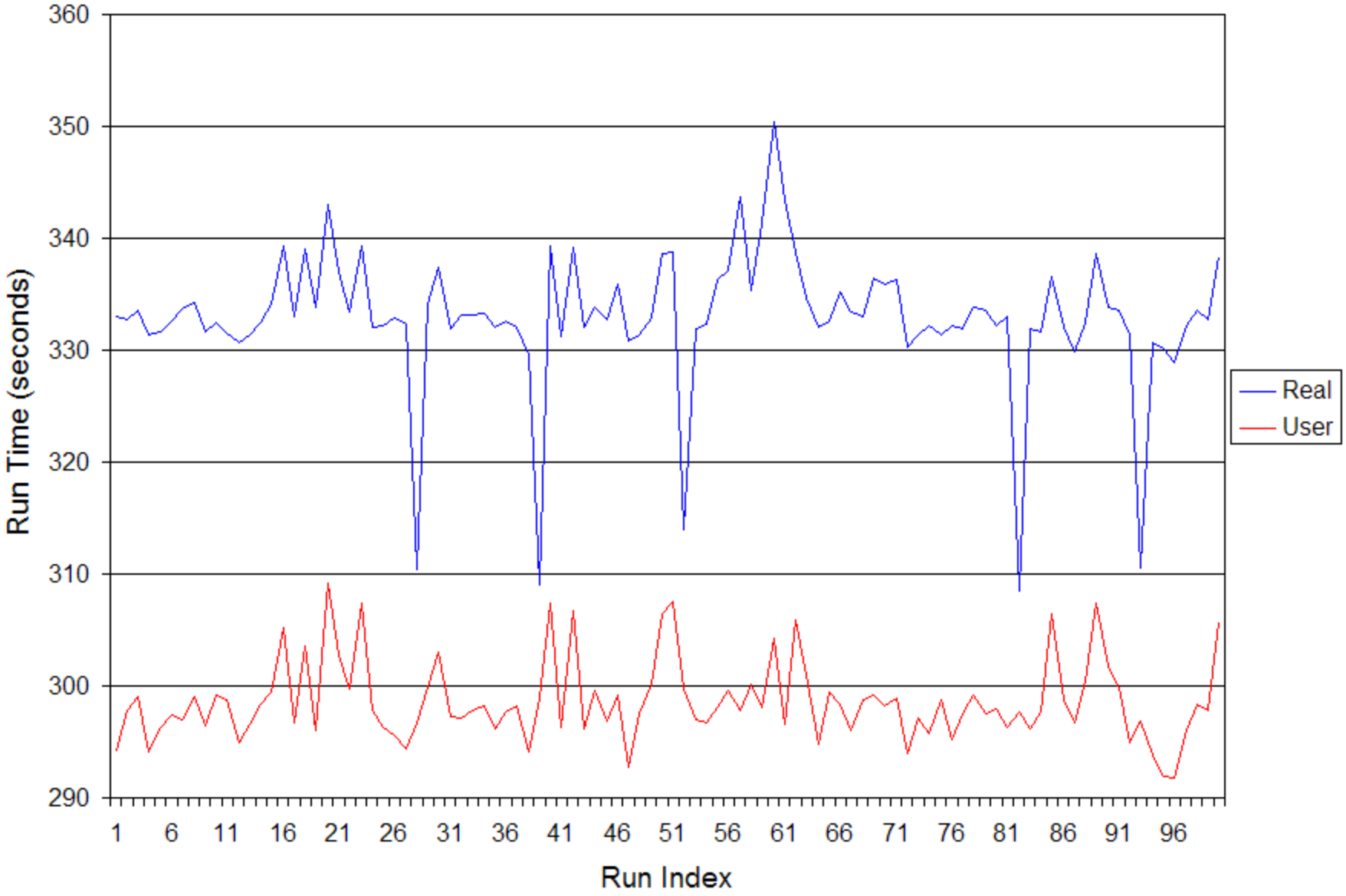
Complete Replication Profile



Validation and Feedback

- Replication Profiles
  - Over 90% concordance with independent biological experiments (FISH)
  - IGF2/H19 Locus (known imprinted) is TNSR
  - Beta Globin (known origin) is a TR50 trough
  - Results used to design additional experiments
    - Probe selection
    - Signal partitioning (select representative samples)
  - Correlation with other genomic markers
    - Histone marks, transcription, gene density, etc

Scaling Up to Whole Genome



Summary

- General framework for profiling DNA replication timing data from discrete replication pools
- High concordance with independent experiments and feedback to experimental design
- Efficient algorithms allow processing of whole human genome data in under 10 hours
  - Parallelization speedup 10-fold
- Data available through world-wide collaboration with NIH ENCODE consortium

Acknowledgements

- Funding:
  - National Institutes of Health (NIH)
    - National Human Genome Research Institute (NHGRI)
- Data Availability:
  - Microarray Data: ArrayExpress E-MEXP-708
  - TR50 & Segregation: UCSC Genome Browser
- Collaborators:
  - Anindya Dutta, Neerja Karnani, Ankit Malhotra, Gabriel Robins

UNIVERSITY of VIRGINIA

Collaboration Avenues

- Molecular Biologists
  - UNO Biology: Dr. Vaniyambadi Sridhar (NSF)
  - RIC: 12 Biology Research Faculty
- Computational Biologists
  - UNO CS: Dr. Dongxiao Zhu (LaBoR)
- Other Complementary Scientists
  - Statisticians (Hardcore Math)
  - Distributed Computing (LONI)