

INTRODUCTION

Genomic structural changes known as copy number alterations (CNAs) have a role in tumor progression. CNAs are chromosomal changes where regions are either amplified or deleted. They can range in length from 100 kb to the entire chromosome. CNAs are known to affect gene function in bladder cancer tumors, and 75,000 new diagnoses of bladder cancer are expected this year in the US alone [1]. Our data consists of measurements from bladder cancer tumor tissue from 93 patients (GSE39281 [1]) whose profiles are in muscle invasive and non-muscle invasive subgroups. It is thought that these subgroups have varying CN profiles that are similar within groups but differ across groups.

GOALS

- Fit the aCGH data using Haar wavelets
- Use functional response regression [2] to characterize CN profiles of muscle invasive and non-muscle invasive patients
- Perform methods on simulated data where the “true” function is known to confirm validity for aCGH data

ARRAY COMPARATIVE GENOMIC HYBRIDIZATION (ACGH)

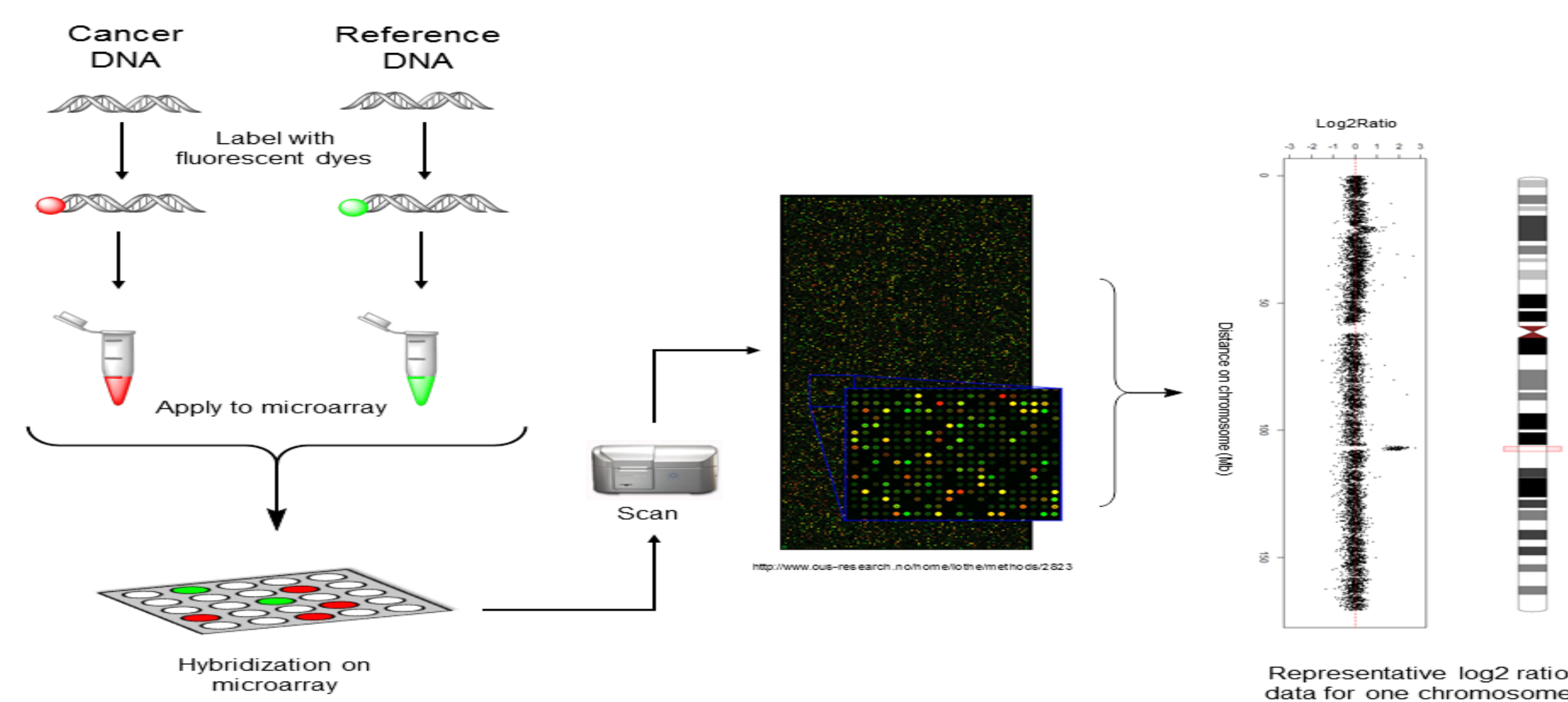


Figure 1: aCGH measures relative DNA quantities by measuring $\log_2$ ratio between test and reference signal [4]. Platform: Agilent-022060 SurePrint G3 Human CGH Microarray 4x180K

- aCGH compares DNA isolated from cancer tissue against a normal reference
- Genomic DNA are fluorescently labeled with Cy3 or Cy5 dye
- Labeled DNA hybridized to probes on array
- A $\log_2$ of the tumor CN to normal CN ratio is reported for each probe

VISUALIZING CHROMOSOME 11 CNAs

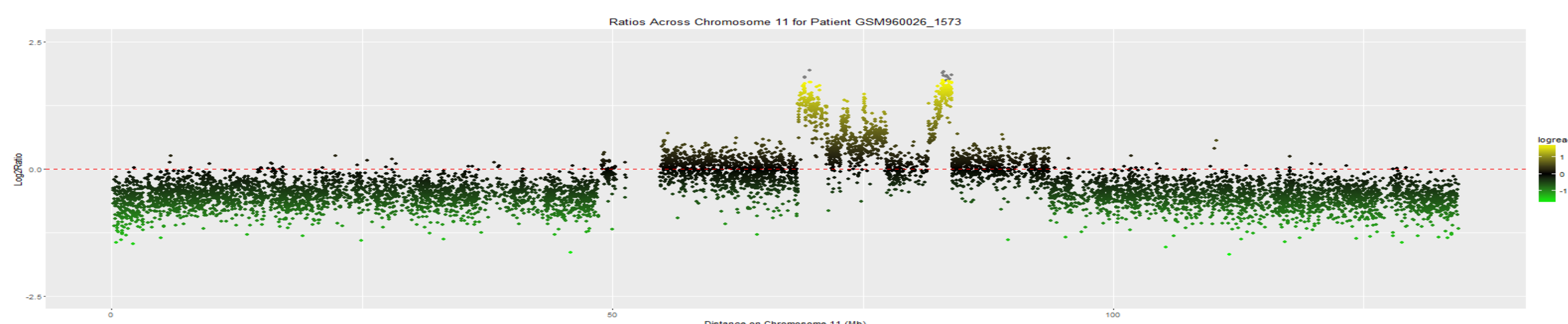


Figure 2: CN readings of chromosome 11 for one patient. X-axis: Position on the chromosome (Mb); centromere at 54Mb. Y-axis: $\log_2$ -ratio of CN readings. Each point is a $\log_2$ ratio of the CN at specified location. Values above dashed zero reference line are amplifications, and values below are deletions.

USING WAVELETS TO MODEL FUNCTIONAL CN PROFILES

- We are viewing CNA profiles as functional data, which are best represented as piecewise constant functions
- Wavelets are basis functions that can be used to represent other functions
- Wavelet properties that are suitable for CN data:
 - Data representation is in location and frequency domains
 - Multiscale, with representations at progressively finer levels of detail
 - Form sparse representation of functions, coefficients of basis functions believed to represent noise are removed by thresholding
- Haar wavelets are well able to capture abrupt changes and discontinuities

The *Haar mother wavelet* is given by:

The *Haar father wavelet* is given by:

$$\psi(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}), \\ -1 & x \in [\frac{1}{2}, 1), \\ 0 & \text{otherwise} \end{cases} \quad \phi(x) = \begin{cases} 1 & x \in [0, 1], \\ 0 & \text{otherwise} \end{cases}$$

The wavelet expansion for a given vector at scale j and level k is given by:

$$f(x) = \sum_k c_{j_0,k} \Phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k d_{j,k} \Psi_{j,k}(x)$$

where $j = j_0, j_0 + 1, \dots$, $c_{j_0,k}$ is the scaling coefficient, $d_{j,k}$ is the detail coefficient, $\phi_{j_0,k}(x)$ is the father wavelet, and $\psi_{j,k}(x)$ is the mother wavelet.

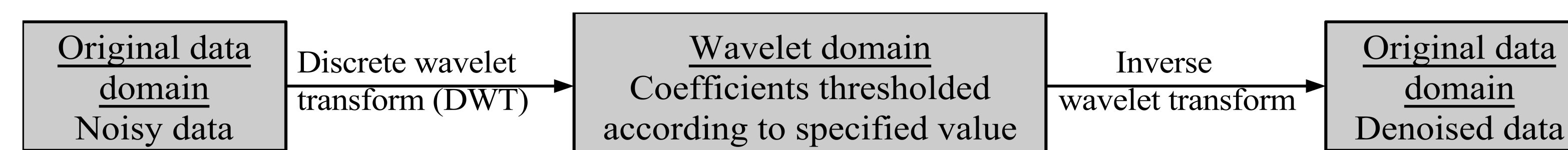


Figure 3: Schematic on how wavelets are used to fit noisy data to create filtered functions

Underlying piecewise, constant CN functions are represented after father and mother wavelet functions act as filters on the $\log_2$ ratio measurements

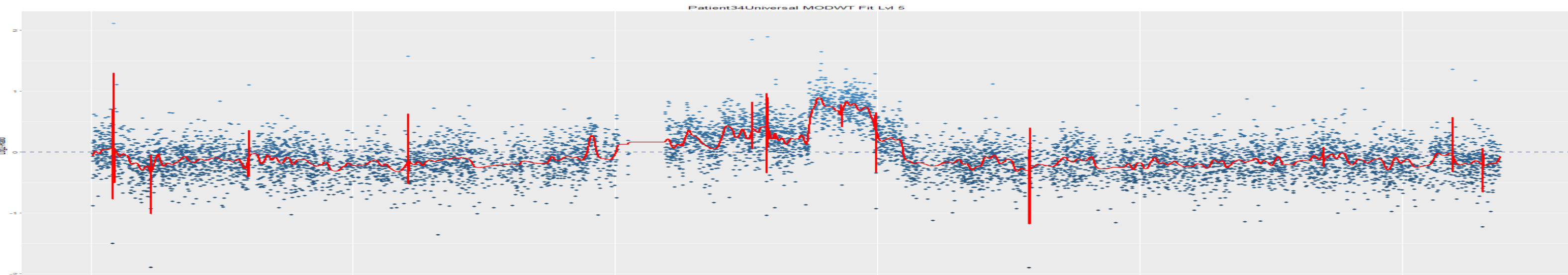


Figure 4: Single patient CN Chromosome 11 profile (blue) with maximal overlap DWT (MODWT) fit overlaid in red

FUNCTIONAL RESPONSE REGRESSION

We're interested in the effect of muscle invasive bladder cancer on the CN profile. Functional response regression allows us to estimate the functional outcome (CN profile) using a categorical predictor (variable indicating muscle invasive or non-muscle invasive). The functional response regression model for p scalar covariates and functional outcome $Y_i(t)$ for individual i is given by:

$$Y_i(t) = \sum_{a=1}^p X_{ia} B_a(t) + E_i(t),$$

$B_a(t)$ is a functional coefficient representing the effect of covariate X on the functional response at location t , and $E_i(t)$ is the error.

ACGH DATA SIMULATIONS

- We developed a simulation method for generating aCGH CN profiles through a series of various probability distributions, which creates a known “true” function
- Noise is added to “true” function to simulate aCGH data with realistic features
- Used to confirm ability of wavelets to capture underlying function from the noisy aCGH data

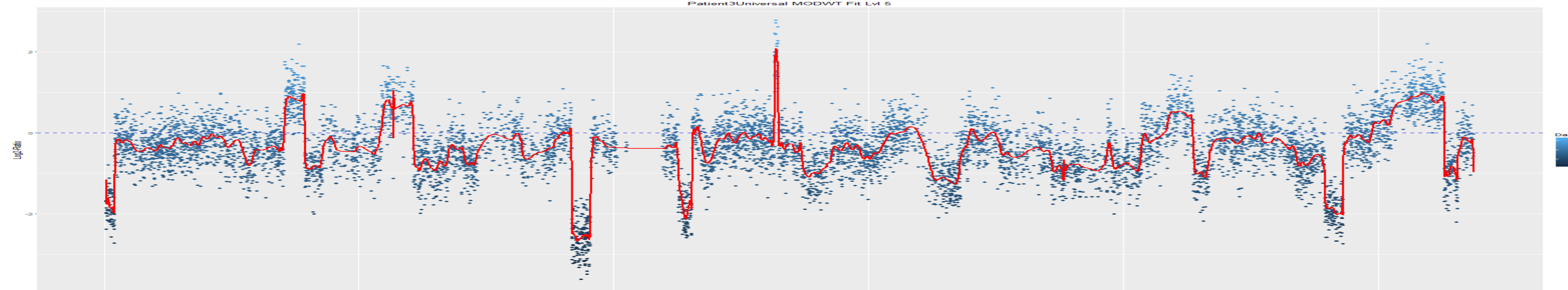


Figure 5: Simulation of noisy aCGH data (blue) with MODWT Universal threshold level 5 fit overlayed in red

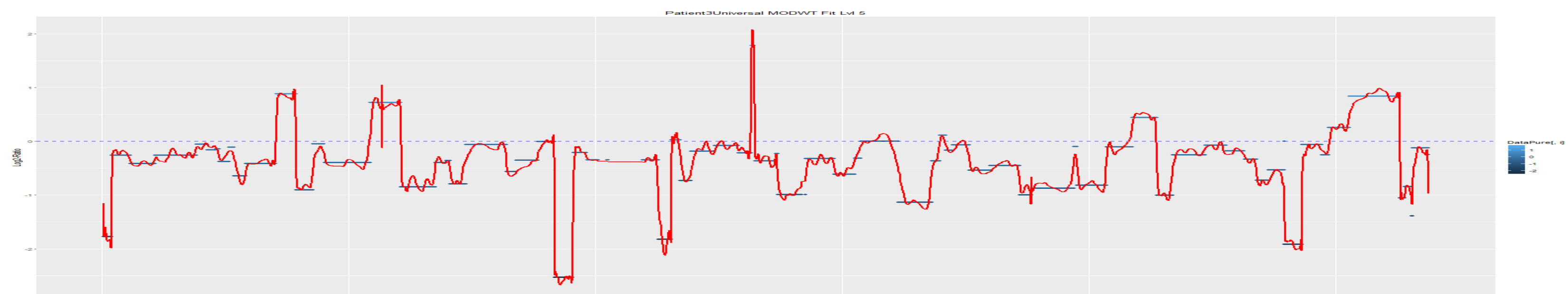


Figure 6: Known true function (blue) of simulated aCGH data with MODWT Universal threshold level 5 fit overlaid in red

RESULTS AND DISCUSSION

Preliminary results show functional response regression is capable of capturing the differences between subgroups. While optimal parameter settings have not been reached, results consistently show three main areas of interest on chromosome 11 in muscle invasive cases. These include bladder cancer-associated genes from 0 to 7 (Mb), IGF2, SMPD1, and RRM1; and CD44 and TRAF6, found between 27 and 37 (Mb). The third region, from 64 to 70 (Mb), is consistently amplified across all patients, and it includes genes RIN1, FG19, FGF4, and ANO1 [5]. We carried out the functional data regression of the CN ratios in the MATLAB based functional regression software WFMM [3].

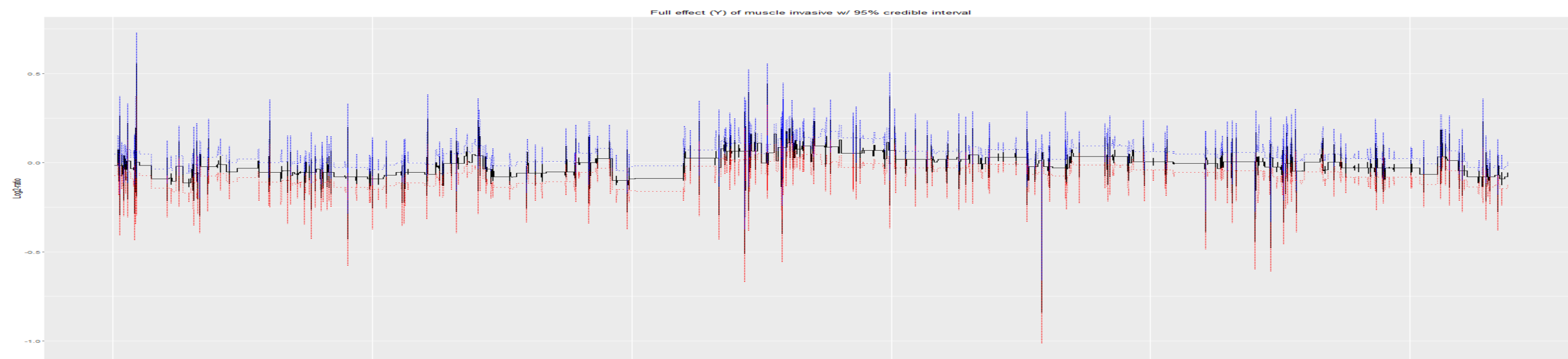


Figure 7: Coefficient function with 95% credible intervals (blue and red) for all profiles in GSE39281, representing main effect on CN of bladder cancer muscle invasiveness for Chromosome 11.

From these analyses, we can infer that functional regression using wavelet basis functions provides an important tool for examining CN profiles in cancer patients.

REFERENCES

- 1 Riester, M, *et al.* (2014) Integrative analysis of 1q23.3 copy-number gain in metastatic urothelial carcinoma. *Clinical Cancer Research*, 20: 1873-1883.
- 2 Morris, JS (2015) Functional regression. *Annual Reviews of Statistics and Its Application*, 2: 321-359.
- 3 Morris, JS, and RJ Carroll (2006) Waveletbased functional mixed models. *Journal of the Royal Statistical Society: Series B*, 68: 179-199.
- 4 Lynch, ML. (2015) Bayesian functional data methods in copy number alteration studies: Applications in urothelial bladder carcinoma. *JSM*.
- 5 Cotterill SJ. Home Page, Cancer Genetics Web: <http://www.cancer-genetics.org/index.htm>