



## Understanding the relationship between cotton fiber properties and non-cellulosic cell wall polysaccharides

Rajasundaram, Dhivyaa; Runavot, Jean-Luc; Guo, Xiaoyuan; Willats, William George Tycho; Meulewaeter, Frank; Selbig, Joachim

*Published in:*  
P L o S One

*DOI:*  
[10.1371/journal.pone.0112168](https://doi.org/10.1371/journal.pone.0112168)

*Publication date:*  
2014

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Rajasundaram, D., Runavot, J-L., Guo, X., Willats, W. G. T., Meulewaeter, F., & Selbig, J. (2014). Understanding the relationship between cotton fiber properties and non-cellulosic cell wall polysaccharides. *P L o S One*, 9(11), [e112168]. <https://doi.org/10.1371/journal.pone.0112168>



# Understanding the Relationship between Cotton Fiber Properties and Non-Cellulosic Cell Wall Polysaccharides

Dhivyaa Rajasundaram<sup>1,2</sup>, Jean-Luc Runavot<sup>3</sup>, Xiaoyuan Guo<sup>4</sup>, William G. T. Willats<sup>4</sup>, Frank Meulewaeter<sup>3</sup>, Joachim Selbig<sup>1,2\*</sup>

**1** Institute of Biochemistry and Biology, University of Potsdam, Potsdam-Golm, 14476, Germany, **2** Max-Planck Institute of Molecular Plant Physiology, Potsdam-Golm, 14476, Germany, **3** Bayer CropScience NV-Innovation Center, Technologiepark 38, 9052 Gent, Belgium, **4** Department of Plant and Environmental Sciences, Faculty of Sciences, University of Copenhagen, Thorvaldsensvej, 40 1.1871, Fredriksberg C, Denmark

## Abstract

A detailed knowledge of cell wall heterogeneity and complexity is crucial for understanding plant growth and development. One key challenge is to establish links between polysaccharide-rich cell walls and their phenotypic characteristics. It is of particular interest for some plant material, like cotton fibers, which are of both biological and industrial importance. To this end, we attempted to study cotton fiber characteristics together with glycan arrays using regression based approaches. Taking advantage of the comprehensive microarray polymer profiling technique (CoMPP), 32 cotton lines from different cotton species were studied. The glycan array was generated by sequential extraction of cell wall polysaccharides from mature cotton fibers and screening samples against eleven extensively characterized cell wall probes. Also, phenotypic characteristics of cotton fibers such as length, strength, elongation and micronaire were measured. The relationship between the two datasets was established in an integrative manner using linear regression methods. In the conducted analysis, we demonstrated the usefulness of regression based approaches in establishing a relationship between glycan measurements and phenotypic traits. In addition, the analysis also identified specific polysaccharides which may play a major role during fiber development for the final fiber characteristics. Three different regression methods identified a negative correlation between micronaire and the xyloglucan and homogalacturonan probes. Moreover, homogalacturonan and callose were shown to be significant predictors for fiber length. The role of these polysaccharides was already pointed out in previous cell wall elongation studies. Additional relationships were predicted for fiber strength and elongation which will need further experimental validation.

**Citation:** Rajasundaram D, Runavot J-L, Guo X, Willats WGT, Meulewaeter F, et al. (2014) Understanding the Relationship between Cotton Fiber Properties and Non-Cellulosic Cell Wall Polysaccharides. PLoS ONE 9(11): e112168. doi:10.1371/journal.pone.0112168

**Editor:** David D. Fang, USDA-ARS-SRRC, United States of America

**Received:** September 17, 2014; **Accepted:** October 6, 2014; **Published:** November 10, 2014

**Copyright:** © 2014 Rajasundaram et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by the European Union Seventh Framework Programme (FP7 2007–2013) under the WallTraC project (grant agreement No. 263916). This paper reflects the authors' views only. The European Community is not liable for any use that may be made of the information contained herein. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Co-authors Dr. Jean-Luc Runavot and Dr. Frank Meulewaeter are employed by Bayer CropScience NV, Innovation Center. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* Email: jselbig@uni-potsdam.de

## Introduction

Cell walls, the key determinant of overall plant growth and development are primarily composed of polysaccharides, namely cellulose, hemicellulose, and pectins, lignin, and structural proteins [1,2]. Cell wall biology has been an area of prominent research over many years with the use of novel technologies to probe these higher order structures in the native state. Since the early 1970's, comparative biochemical analyses revealed that all plant cell walls share several common features. However, they exhibit diversity with respect to their chemical composition [3–5]. Indeed, cell walls are structurally complex as they are constantly remodeled and reconstructed during plant growth and development. Also, walls are modulated according to functional requirements, thereby limiting our knowledge on cell wall design [6–8]. Biochemical analyses complemented by genetic analyses have identified genes and gene products associated with cell wall synthesis. However, an understanding of how these genes are expressed across cells of

different tissues and their impact on cell wall design and maintenance is still lacking [9–11]. Furthermore, the glycan-rich cell walls influence the nutritional and processing properties of plant based products such as pulp for paper manufacture, textile fibers, timber products, pharmaceuticals, and materials for fuel and composite manufacture [12–15]. Therefore, understanding the plant cell walls is not only fundamental to plant sciences but also of industrial relevance.

Microarrays are widely used in plant research for the high throughput analysis of nucleotides, proteins and increasingly, carbohydrate [16–18]. Carbohydrate microarrays also referred to as glycan arrays enable hundreds of glycans to be analyzed in parallel. Glycans on the arrays can include oligosaccharides, polysaccharides, glycoproteins and glycolipids [19–21]. Glycan arrays have several biological and medical applications which include glycoproteomic methods to identify new glycoproteins and glycans [22,23], characterization of glycan probes [24], profiling carbohydrate-lectin interactions [25,26], glycosaminoglycans-

growth factor and cytokine interactions [27,28], pathogen-induced antibody interaction [29,30], cancer-antibody induced interaction [31,32], carbohydrate-virus interactions [33], quantitative carbohydrate-protein interactions [34], and drug discovery [35,36].

Comprehensive microarray polymer profiling (CoMPP), a microarray based glycan screening method is mostly used for high throughput characterization of plant cell walls. In this technique, generation of microarrays by sequential extraction of cell wall polysaccharides and screening samples against a large number of well-defined cell wall probes such as antibodies, carbohydrate binding proteins and modules is done. This methodology was first described in *Arabidopsis thaliana* and *Physcomitrella patens* [37]. In the study of Singh et al, application of CoMPP to study cotton fibers showed that towards the end of elongation, there was a loss in certain cell wall polymer epitopes [38]. Despite the availability of glycan arrays from several experiments, computational analysis has mostly been restricted to collection of glycobiology information in databases, motif analysis of glycans, and oligosaccharide structure determination [39–41].

In our study, we used the glycan array technology to study cotton fibers, one of the most important raw materials for the textile industry. There are four different domesticated species producing cotton fibers namely *Gossypium hirsutum* ('Upland cotton'), *Gossypium barbadense* ('Pima' or 'Egyptian' cotton), *Gossypium arboreum* ('Tree cotton'), and *Gossypium herbaceum* [42]. The development of cotton fibers occurs in four major stages: initiation, elongation, secondary wall synthesis and maturation. Although much work has already been done on the cotton fiber transcriptome, the key question in cotton fiber research is to link the cell wall profile of different cotton types to the cotton fiber properties and to a better understanding of fiber development [43–46]. Here, we aim to study the relation between fiber properties and non-cellulosic polysaccharide composition using univariate and multivariate regression based approaches on a diverse set of cotton fibers. To this end, we analyzed two datasets for the same cotton fibers: a glycan array profile and the physical fiber properties as determined by HVI and AFIS. We elucidated the usefulness of regression based approaches to determine the functional relationship between the two datasets and we also selected a subset of variables which have a good prediction of the phenotypic traits.

## Materials and Methods

### Plant material and evaluation of phenotypic traits

In this study, we used 32 different cotton lines of which three are from *Gossypium arboreum*, three from *Gossypium barbadense*, two from *Gossypium herbaceum* and 24 from *Gossypium hirsutum*. The cotton lines used in this study are listed in Table S1, including the plant introduction number (PI number) from the USDA National Plant Germplasm System (<http://www.ars-grin.gov/npgs/>). Seeds were sown in soil compost and plants were grown at constant conditions in a greenhouse at 26–28°C during a 16 h photoperiod. Mature cotton fibers were collected by harvesting all fully open bolls from several plants. The impact of boll position and plant-to-plant variation was minimized by mixing the fiber from all harvested bolls. Two types of analyses were performed on these fibers, the first being the glycan array measurements (Table S2) and the second being fiber characteristics/phenotype measurements (Table S1). For each line, High Volume Instrument (HVI) and Advanced Fiber Information System (AFIS) measurements were performed on 40 g of mature cotton fiber by CIRAD

(France) according to the standard methods ASTM D3818-92 and D5867-05. These measurements were done on 6 and 5 replicates for HVI and AFIS, respectively, except for micronaire where only 2 replicates were performed. Five fiber characteristics which include length from HVI and AFIS, strength, elongation and micronaire were selected for further analysis due to their importance for textile processing. Length HVI refers to the average fiber length of the longer 50% of fibers in a given sample. Length AFIS (W) L deduces length parameters from individual fiber measurements. Strength of the cotton fiber refers to the force required to break a bundle of fibers 1 tex in size (1 tex equals the weight in grams of 1000 meters of fibers). Elongation of the cotton fibers is the measurement of the elasticity of cotton fibers with a higher number indicating more elasticity. Micronaire is obtained by measuring the resistance of the fibers to airflow and depends on the fiber fineness and degree of maturation.

### Comprehensive Microarray Polymer Profiling (CoMPP) of mature cotton fiber cell wall

CoMPP analysis was performed on mature cotton fibers as previously described by [38] with minor modifications. Mature cotton fiber samples were extracted sequentially in 50 mM cyclohexanediamine tetraacetic acid (CDTA) and 4 M Sodium hydroxide (NaOH) with 1% (v/v) sodium tetrahydridoborate (NaBH<sub>4</sub>). These two solvents were used to extract pectins and non-cellulosic polysaccharides, respectively. For each line, 300  $\mu$ l of solvent was added to 10 mg of sample and incubated with shaking for 2 h. After centrifugation, supernatant from each extraction was printed in four replicates and four dilutions (1:2, 1:6, 1:18 and 1:54 [v/v] dilutions). Cadoxen extraction was omitted because it is mainly used to extract cellulose which we do not aim to analyse in our study. The array was probed with eleven monoclonal antibodies (mAbs) recognizing different carbohydrate epitopes as listed out in Table 1. A heat map was generated to display the relative intensity of each signal to the maximum signal observed within each antibody detection (Table S2). CoMPP is a semi-quantitative technique and should not be taken to obtain absolute amounts. Practically speaking, we set the maximum value in the whole data sheet as 100 and the other values are divided by this maximum value and multiplied by 100 to obtain numbers comprised between 0 and 100. When the quantification is done, the arrays are manually checked to make sure that there are clear dots on it and not only background or noise. The negative control is an array incubated with 5% milk in PBS and probed with secondary antibody and then developed as the others.

### Pre-processing of the data

For the statistical analysis, we use R version 3.1.0 on a 64 bit linux platform [56]. The numerical values from both datasets were of different physical quantities and on different scales of magnitude. Moreover, there is no external knowledge that variables with higher numeric variation should be considered more important. Standardization of the raw data was done by computing z-scores of the raw data. Z-scores were calculated for each data point by subtracting the mean and dividing by the standard deviation of all data points.

### Linear methods to delineate the relationship between the two datasets

Multiple regression models the relationship between a single scalar response variable and a set of explanatory (or independent) variables. Here, we used multiple regression analysis to model which of the cell wall probes were associated to the fiber

**Table 1.** List of probes used in the glycan array.

Probes used in the analysis	Specificity of the probes	Reference	Source
BS-400-2	(1,3)- $\beta$ -D-glucan (callose)	[47]	Purchased from Biosupplies (Australia)
JIM5	Partially methyl-esterified homogalacturonan (HG)	[48]	Paul Knox lab
LM19	Un-esterified homogalacturonan (HG)	[49]	Paul Knox lab
JIM13	Arabinogalactan (AGP)	[50]	Paul Knox lab
JIM20	Extensin glycoproteins	[51]	Paul Knox lab
LM11	Xylan	[52]	Paul Knox lab
LM15	XXXG xyloglucans (XG) epitope	[53]	Paul Knox lab
LM24	XXLG and XLLG xyloglucan (XG) epitopes	[24]	Paul Knox lab
LM25	XXLG and XLLG xyloglucan (XG) epitopes	[24]	Paul Knox lab
BS-400-4	Mannan	[54]	Purchased from Biosupplies (Australia)
LM21	Mannan	[55]	Paul Knox lab

doi:10.1371/journal.pone.0112168.t001

characteristics. This allowed us to determine the overall fit (variance explained) of the model and the relative contribution of each of the cell wall probes to the total variance explained. The results from the analysis were reported in the coefficients and ANOVA tables. Summary of the fitted model object gave an account of the residuals, the estimates of the intercept, the slope (with the results of a t-test), the residual standard error, the  $R^2$  statistic and the results of an F-test. The terms used in the output of regression analysis are defined as follows: residual standard error is the standard deviation of the data about the regression line. The squared multiple correlation coefficient ( $R^2$ ) is the proportion of variability in the response that is fitted in the model and the F value is a test statistic to decide whether the model as a whole has statistically significant predictive capability. p values give the statistically significant predictive capability in the presence of other variables [57,58]. Based on this, five models were selected to determine which of the cell wall polysaccharides play an important role in determining that particular fiber characteristic.

In addition to the multiple regression analysis, relationships between multiple dependent and independent variables were investigated simultaneously using canonical correlation analysis (CCA). The two sets of data were represented by matrices X (dimension  $n \times p$ ) and Y (dimension  $n \times q$ ) and columns in X and Y denote the variables p (glycan measurements) and q (fiber characteristics) respectively. Classification of variables as dependent or independent is of little importance for the statistical estimation of the canonical functions as canonical correlation finds linear combinations of sets of multiple dependent and independent variables which are maximally correlated [59,60].

The first step in CCA was to derive one or more canonical function between the glycan and phenotypic measurements. Each function consisted of a pair of variates, one representing the cell wall probes and the other representing the fiber characteristics. The maximum number of canonical variates (functions) that could be extracted from the sets of variables equals the number of variables in the smallest data set, independent or dependent. As a result, the first pair of canonical variates was derived so as to have the highest intercorrelation possible between the glycan array and the fiber measurements. Technically, the second pair of canonical variates exhibits the maximum relationship between the two sets of variables (variates) not accounted for by the first pair of variates and successive pairs of canonical variates were based on residual variance. Therefore each of the pairs of variates is orthogonal and independent of all other variates derived from the same set of data.

The strength of the relationship between the pairs of variates obtained from both datasets was determined by the canonical correlation. An estimate of shared variance between the canonical variates was provided by the squared canonical correlations, also called canonical roots or eigenvalues. The statistical significance of each canonical function was assessed using multivariate tests of significance namely Wilk's lambda, Hotelling's trace, Pillai's trace and Roy's greatest characteristic criterion (Roy's gcr). The statistically significant canonical functions were then interpreted using canonical loadings, cross-loadings and redundancy index [61–65]. We used the “mixOmics” package [66] in R to perform the canonical correlation analysis.

### Sparse partial least square regression to predict the cell wall probes associated to fiber characteristics

Partial least squares (PLS), a well-known regression technique dealing with collinear matrices, clearly has an edge over other regression techniques [67]. Unlike CCA, the PLS latent variables are linear combinations of the variables based on the maximization of covariance but do not allow feature selection. There are many variants of PLS of which we focused on a sparse partial least squares approach (sPLS) which includes a built-in feature to select variables while integrating the data. We used the “mixOmics” package [66] in the regression mode. Specifically, we use a two block data setup, X be the  $n \times p$  matrix and Y be the  $n \times q$  matrix where n denotes the samples, variables p and q denote the glycan measurements and fiber characteristics respectively. Sparse PLS, based on lasso regression penalizes the loading vectors using singular value decomposition and has an additional advantage to perform better even when the covariates are highly correlated. We used sPLS in the regression mode and the aim was to model the relationship between the variables and also predict one group of variables from the other [68–71].

## Results

### Standardization of the raw data

In this study, we attempted to assess the relationship between the cell wall polysaccharides and the physical fiber properties of mature cotton fibers, the data of which are provided as Table S1 and S2. The glycan array values used for the regression analysis were the sums from the CDTA and the NaOH extractions as performing the analysis using the individual values gave the same

correlations. For the fiber characteristics dataset, the values were in different units and scales such as mm (for length), g/tex (for strength), and percentage (for elongation). To make the fiber characteristics dataset compliant to the glycan array, the raw data were jointly standardized using z scores prior to the analysis.

### Modelling the fiber properties using linear regression models

We investigated the linear relationship between the fiber properties and their corresponding array values by a series of regression analyses. Multiple regression models were built considering one fiber characteristic at a time as the dependent variable and multiple probes as the independent variables. Five such models were predicted for the phenotypic traits and the overall model prediction result (Table 2) shows that the model for length HVI, length AFIS and micronaire are statistically significant. The significant predictor variables of length HVI are BS-400-2, LM19 and the ones for length AFIS include BS-400-2, JIM5, JIM20, LM15, and LM19. LM15, LM19, LM24 and LM25 are the significant predictor variables for the model predicting cotton fiber micronaire and the overall model has a p value of 4.906e-06. The models for strength and elongation do not show any statistical significance.

### Assessing the relationship between multiple probes and all of the fiber characteristics simultaneously using canonical correlation analysis

The multiple regression analysis can predict the value of a single (metric) dependent variable from a linear function of a set of independent variables. However, to explore the relationship of sets of multiple predictor variables (probe measurements) to sets of multiple response variables (phenotypic traits) CCA was used. As CCA uses information from all the variables in both the predictor and response sets, it serves as a more efficient approach than methods routinely used, such as multiple linear regression.

For the CCA analysis, the glycan array measurements (probed by 11 antibodies) are designated as the set of independent variables. The fiber characteristics namely length AFIS, length HVI, strength, elongation and micronaire were specified as the set of dependent variables (Figure 1). However, it is of little importance to classify the variables as independent or dependent as the technique aims to maximize the correlation between the two sets of variables. In Figure 1, the terms  $r_{x1}$  to  $r_{x11}$  represent the canonical loadings which reflect the variance that the eleven variables from the glycan array shares with the independent canonical variate  $U_1$ . Similarly the terms  $r_{y1}$  to  $r_{y5}$  represent the canonical loadings which reflect the variance that the five

phenotypic variables share with the dependent canonical variate  $V_1$ . The canonical correlation between the independent and dependent canonical variates is measured by the canonical functions which are represented by  $R^2_{c1}$  to  $R^2_{c5}$ . The statistical problem involved identifying any latent relationships (relationships between composites of variables rather than the individual variables themselves) between the glycan and the fiber measurements.

The canonical correlation which is based on the linear relationship of the glycan array data and fiber characteristics was computed to derive five canonical functions. Each of these functions consists of a pair of variates, one for the glycan array data and the other for the fiber characteristics. Since the study includes 11 independent variables and 5 dependent variables, the maximum number of canonical functions which could be derived is five (Table 3).

In addition to tests of each canonical function separately, multivariate tests of these five functions simultaneously were also performed. The test statistics employed include Wilks' lambda, Pillai's criterion, Hotelling's trace, and Roy's gcr. Table 4 details the p-values from the multivariate test statistics, which all indicate that only the first canonical function, taken collectively, is statistically significant at 1% level.

From the results of these tests, we proceeded to interpret other aspects of the analysis based on the first canonical function. A redundancy index was calculated for the independent and dependent variates of the first function in Table 5. The redundancy index is calculated as the average loading squared times the canonical  $R^2$ . As can be seen, the redundancy index for the dependent (0.191) and independent variates (0.200) is quite low. The low values result from the relatively low shared variance in the dependent variates (0.214) and independent variates (0.225), not the canonical  $R^2$ . With such a small percentage, this is an example of a statistically significant canonical function that does not have practical significance because it does not explain a large proportion of the dependent variables' variance.

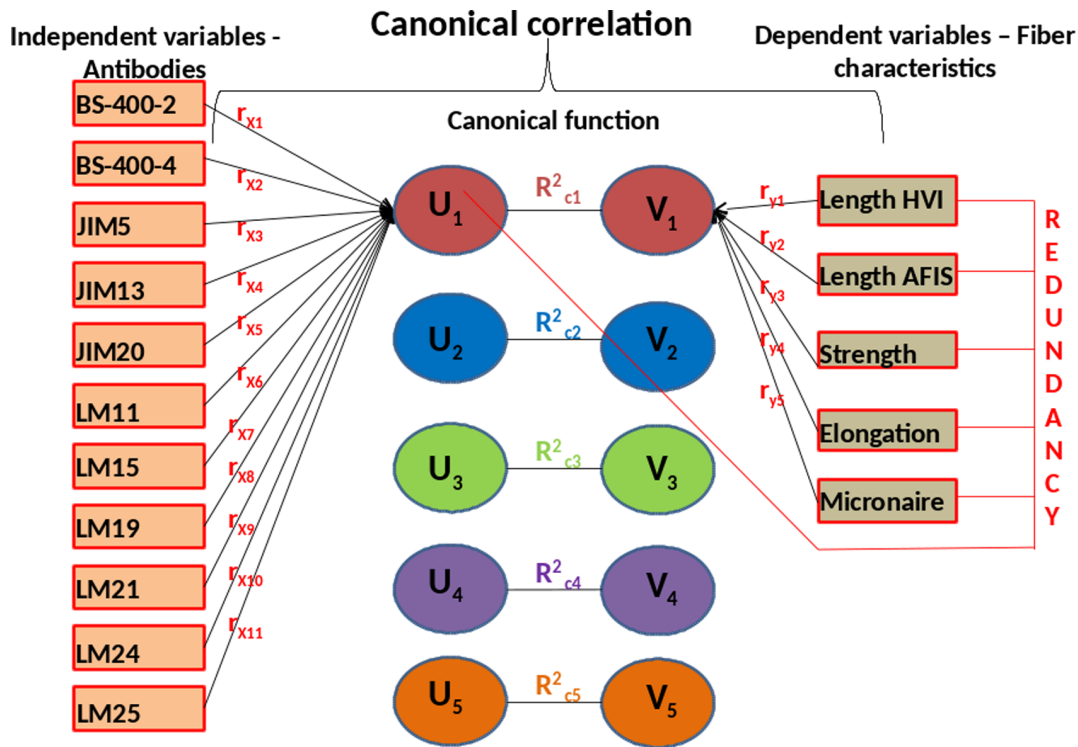
The interpretations involve examining the canonical functions to determine the relative importance of each of the original variables in deriving the canonical relationships (Table 6). The three methods for interpretation are (1) canonical weights (standardized coefficients), (2) canonical loadings (structure correlations), and (3) canonical cross-loadings.

Table 6 contains the standardized canonical weights for each canonical variate for both dependent and independent variables. As mentioned earlier, the magnitude of the weights represent their relative contribution to the variate. Based on the size of the weights, the order of contribution of independent variables to the

**Table 2.** Summary statistics of the five possible multiple regression models.

Fiber characteristics	Residual standard error	Multiple R-squared	Adjusted R-squared	F-statistic	p-value	Significant predictors
Length HVI	0.696	0.706	0.545	4.372 on 11 and 20 DF	0.002	BS-400-2, LM19
Length AFIS	0.632	0.720	0.566	4.677 on 11 and 20 DF	0.001	BS-400-2, JIM5, JIM20, LM15, LM19
Strength	0.940	0.378	0.036	1.107 on 11 and 20 DF	0.404	JIM20
Elongation	0.825	0.376	0.033	1.098 on 11 and 20 DF	0.410	–
Micronaire	0.469	0.851	0.769	10.4 on 11 and 20 DF	4.906e-06	LM15, LM19, LM24, LM25

doi:10.1371/journal.pone.0112168.t002



**Figure 1. Canonical correlation analysis maximizes the correlation between the linear combination of the cell wall polysaccharides in the glycan array and the fiber properties.** In this figure, given a linear combination of  $X$  variables:  $U_1 = f_{1x1} + f_{2x2} + \dots + f_{px}X_p$  and a linear combination of  $Y$  variables:  $V_1 = g_1Y_1 + g_2Y_2 + \dots + g_qY_q$ , the first canonical correlation is the maximum correlation coefficient between  $U_1$  and  $V_1$ , for all  $U_i$  and  $V_i$ .  
doi:10.1371/journal.pone.0112168.g001

first variate is LM19, LM25, JIM5, LM15, BS-400-4, LM21, LM24, JIM13, and JIM20 and the dependent variable order on the first variate is micronaire followed by length AFIS, length HVI, strength and elongation. Because canonical weights are typically unstable, particularly in instances of multicollinearity, owing to their calculation solely to optimize the canonical correlation, the canonical loading and cross-loadings are considered more appropriate.

Table 6 also contains the canonical loadings for the dependent and independent variates for the first canonical functions. In the first dependent variates, all the five variables had different values of loadings resulting in low shared variance (0.214). This indicates a low degree of inter-correlation among the five dependent variables. Observing the independent variates, there is a different pattern and loading values ranged from 0.06 to 0.77. The variables with the highest loadings on the independent variate are LM25, LM19, LM15, and JIM5. We also observed some loadings

with negative values which include those of BS-400-4, JIM20, and LM11.

In case of the cross loadings, micronaire has a value of  $-0.890$  and interestingly has a negative loading. Length AFIS to some extent has a loading value of  $0.387$  while those of the other variables is low. By squaring these terms, we find the percentage of the variance for each of the variables explained by function 1. The results show that 79.21 percent of the variance in micronaire, 14.97 percent of the variance in length AFIS is explained by function 1 whereas strength, elongation and length HVI have very low values. Similarly for the independent variables' cross loadings, variables LM25, LM19, LM15, JIM5 have high correlations of  $0.73$ ,  $0.67$ ,  $0.61$ , and  $0.61$  respectively. From this information, approximately 51.8% of the variance in LM25, 45.1% of the variance in LM19, 36.3% of the variance in LM15, and 35.7% of the variance in JIM5 is explained by the dependent canonical variates.

**Table 3. Canonical Correlation analysis relating probe signals and fiber characteristics with the measure of overall model fit.**

Canonical function	Canonical correlation	Canonical R <sup>2</sup>	F statistics	p-value
1	0.945	0.883	2.85	1.57338e-05
2	0.868	0.753	1.88	1.072479e-02
3	0.803	0.645	1.30	2.035076e-01
4	0.523	0.273	0.59	8.715784e-01
5	0.342	0.116	0.34	9.040993e-01

doi:10.1371/journal.pone.0112168.t003

**Table 4.** Multivariate tests of significance for the canonical functions.

Canonical function	Wilks' Lambda, using F-approximation (Rao's F):	Hotelling-Lawley Trace, using F-approximation:	Pillai-Bartlett Trace using F-approximation	Roy's largest root using F-approximation
1	1.57338e-05	2.666759e-07	0.00	8.732348e-12
2	1.072479e-02	1.221924e-03	0.042	
3	2.035076e-01	5.475660e-02	0.285	
4	8.715784e-01	8.365373e-01	0.801	
5	9.040993e-01	8.855773e-01	0.848	

doi:10.1371/journal.pone.0112168.t004

The final step of interpretation is examining the signs of the cross-loadings. Examining the signs of the independent variables' cross loadings, those with high correlations have a positive direct relationship whereas BS-400-4, JIM20 and LM11 have an inverse relationship. The four highest cross-loadings of the first independent variate correspond to the variables with the highest canonical loadings as well. Observing the cross loadings of the dependent variables, we see that micronaire has the highest canonical loading and an inverse relationship. Also, elongation is observed to have an inverse relation but since it is of very low value, it was not taken into account.

#### sPLS approach to predict specific cell wall polysaccharides involved in fiber properties

sPLS was computed in the regression mode and the input for the analysis included the 11 cell wall probes along with the five fiber characteristics. The number of dimensions  $H$  to be retained was estimated with the  $Q_h^2$  criterion, for which a value below the threshold 0.0975 indicates a significant contribution for the prediction purpose. The  $Q_h^2$  values calculated for each dimension of the sPLS showed that 2 dimensions were enough to capture the whole information. From Figure 2, we can interpret the results from the sPLS via the correlation circle plot where the predictor variables are in red and the response variables are represented in blue. A correlation circle plot is a graphic tool to represent variables of two different data-types and examine the relationships between the variables and variates. In this plot, variables namely cell wall probes and fiber measurements can be represented as vectors. The relationship between these two data-types is approximated by the inner product between the associated vectors which is defined as the product of the two vector lengths and their

cosine angle. For better interpretation, two circles of radii 0.5 and 1 are represented to visualize the variables. The longer the distance to the origin, the stronger is the relationship between the variables.

Using the interpretation which is detailed, we find that BS-400-4, LM21, and JIM13 share a positive relationship with elongation characteristic of cotton fibers. We were also able to attribute the strength of the cotton fibers to JIM20, LM11 and LM24. Interestingly, LM19, JIM5, LM15, and LM25 were projected diametrically opposite to that of the micronaire in the correlation circle, thereby indicating a strongly negative relationship. Length HVI and length AFIS share a negative relation to BS-400-2. To estimate the significance of the predicted relationships, the root mean squared error prediction (RMSEP) values were computed for each response variable (fiber properties) and ranked according to the absolute value of their loadings in  $v_2$ . The lower the RMSEP value, the better the prediction of the model is. In this case, the model for micronaire was the best one (RMSEP of 0.71), followed by that of length AFIS (1.13), strength (1.14), elongation (1.15), and length HVI (1.21).

Figure 3 displays the graphical representation of the cotton lines in dimension 1 and 2. This plot shows that some of the lines are clustered together, with Acala SJ1, Germaines Acala (GC 352 and GC 362), TAM-90C-19 S, and FM966 forming one cluster, Acala red okra, okra leaf, multiple marker, Tidewater, and TTU 202-1107B forming a second cluster and PIMAS7, Lankart 57, IV4F-91057, GA161, Ting tao tzu ching chung mien, Brymer brown, Malla guza, Selection of SHIH, China 10, Texas rust brown, Tex 1000 and 30834 (A1660) forming a third cluster. Strikingly, some of these clusters contain lines from different *Gossypium* species and lines from one species often belong to multiple clusters. The

**Table 5.** Redundancy analysis of dependent and independent variates for the first canonical function.

Standardized variance of the dependent variables explained by				
Their own Canonical variates (shared variance)			The opposite canonical variates (Redundancy)	
Percentage	Cumulative percentage	Canonical $R^2$	Percentage	Cumulative percentage
0.214	0.214	0.883	0.191	0.191
Standardized variance of the independent variables explained by				
Their own Canonical variates (shared variance)			The opposite canonical variates (Redundancy)	
Percentage	Cumulative percentage	Canonical $R^2$	Percentage	Cumulative percentage
0.225	0.225	0.883	0.200	0.200

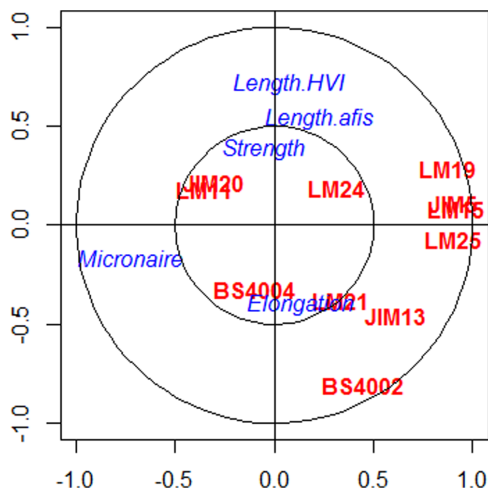
doi:10.1371/journal.pone.0112168.t005

**Table 6.** Canonical weights, loadings, and cross-loadings for the first canonical function.

	Canonical weights	Canonical loadings	Canonical cross-loadings
<b>Dependent variables</b>			
Length HVI	-0.636	0.127	0.120
Strength	-0.226	0.040	0.038
Elongation	-0.033	0.056	0.053
Micronaire	-0.843	-0.941	-0.890
Length AFIS	0.810	0.409	0.387
<b>Independent variables</b>			
BS-400-2	0.184	0.362	0.342
BS-400-4	-0.487	-0.119	-0.113
JIM5	-0.823	0.632	0.598
JIM13	-0.290	0.288	0.272
JIM20	-0.204	-0.376	-0.355
LM11	-0.114	-0.360	-0.340
LM15	-0.719	0.638	0.603
LM19	1.243	0.712	0.672
LM21	0.356	0.275	0.260
LM24	-0.324	0.066	0.062
LM25	1.082	0.767	0.724

doi:10.1371/journal.pone.0112168.t006

variation in fiber characteristics and composition is thus clearly not species-specific. However, one should be careful in interpreting the results from the individual lines as the study was designed to discover correlations between fiber properties and composition and not to study properties of individual lines.



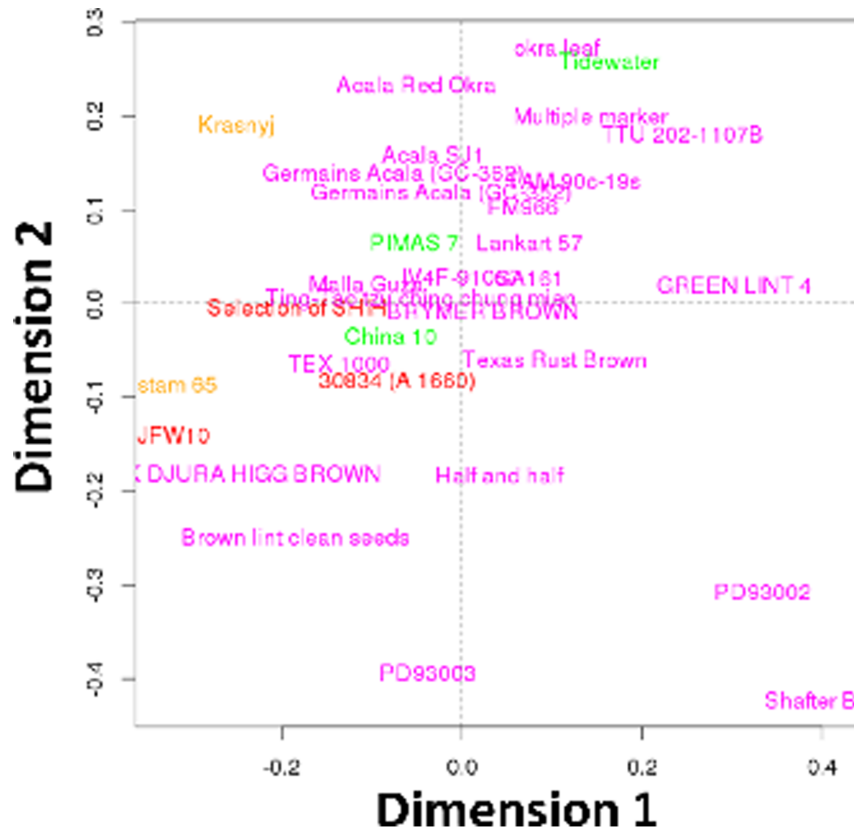
**Figure 2. Graphical representation of the variables selected by SPLS on the first two dimensions predicts specific cell wall polysaccharides linked to the fiber properties.** The coordinates of each variable are obtained by computing the correlation between the latent variable vectors and the original dataset. The selected variables are then projected onto correlation circles where highly correlated variables cluster together. These graphics help to identify association between the two datasets. The correlation between two variables is positive if the angle is sharp  $\cos(\alpha) > 0$ , negative if the angle is obtuse  $\cos(\theta) < 0$ , and null if the vectors are perpendicular  $\cos(\beta) \sim 0$ . doi:10.1371/journal.pone.0112168.g002

## Discussion

Understanding the genetics and physiology of cotton fibers is of importance to the textile industry. There have been numerous studies, both profiling and sequencing based experiments to study cotton fiber development at the transcriptional level. The high degree of transcriptional complexity in the development of cotton fibers has been the focus of these studies [43,72–75]. We used the CoMPP technique in our analysis to study directly the glycan composition of cotton lines from different species. The work presented here demonstrates the potential of glycan microarrays in combination with multivariate statistical approaches for understanding the cell wall composition responsible for the fiber characteristics. Specifically, the use of regression based approaches in our study helps to predict models for each of the fiber trait under study.

We studied the association between glycan array measurements and their relation with fiber characteristics using linear approaches like multiple regression, CCA and sPLS. From the results of multiple regression (Table 2), we were able to predict three models for length HVI, length AFIS and micronaire of cotton fibers but not for strength and elongation characteristics. Moreover, to extend our understanding of the data to situations involving more than one fiber characteristic at a time, CCA was used as it simultaneously models effects of multiple independent variables on multiple dependent variables. As CCA uses information from all the variables in both the exposure and outcome variable sets and maximizes the estimation of the relationship between the two sets, CCA may offer a more efficient approach for assessing the relationship of the cell wall probes with fiber characteristics than methods routinely used such as multiple linear regression. CCA starts with simultaneous consideration of both glycan array measurements and the phenotype measures, limiting the inefficiencies that may accompany conventional multiple testing, and thus, reducing type-1 error. The resulting procedure gives a global view of association between indicators of both datasets. Thus,





**Figure 3. Graphical representation of the cotton lines on the first two sPLS dimensions shows the trend in clustering of specific cotton lines across different species.** Four different species of cotton are shown in different colors. *Gossypium hirsutum* is colored in magenta, *Gossypium barbadense* in green, *Gossypium herbaceum* in orange and *Gossypium arboreum* in red. doi:10.1371/journal.pone.0112168.g003

CCA could be used as a comprehensive approach to extract information from data simultaneously. Another major advantage of using the CCA to multiple regression analysis is to deal with the issue of multicollinearity. In multiple regression, the interpretation is usually based on the significance of weights, which is highly influenced by multicollinearity. If two variables have a high correlation one of them will be completely eliminated even if both have a high correlation to the outcome. In our analysis this is illustrated by JIM5 and LM19 (both detecting homogalacturonan), with both showing a high correlation with micronaire in CCA but only LM19 being identified as a predictor of micronaire in the linear regression model. From the results of the CCA, we obtained an overall picture of associations between the glycan and phenotype measurements, with information about the relative contribution of the variables to that particular canonical variate through canonical loadings. The canonical analysis revealed that the canonical correlation was statistically significant at 1%. However, canonical correlation based methods are statistically difficult to assess as they do not fit into a regression framework. In this context penalized CCA adapted with elastic net (CCA-EN) could be used but the elastic net is similar to a lasso soft-thresholding penalization and the algorithm uses partial least squares and not canonical correlation computations [63]. From [63] it is evident that sPLS made a good compromise between all of these approaches and includes variable selection. Additionally, we used the sPLS approach to be able to predict specific cell wall polysaccharides linked with fiber characteristics. Moreover, sPLS

maximizes the covariance between the latent variables whereas the canonical correlation based methods maximize the correlation.

There were both unique and common findings from the three types of regression analysis. The major and most significant finding in common to all these analyses is that micronaire is negatively correlated with the xyloglucan (XG) and homogalacturonan (HG) probes. One possible explanation for this observation is that cotton fiber with a high micronaire usually has a very thick secondary cell wall resulting in very high levels of cellulose and lower levels of the non-cellulosic components. However, we do not find a negative correlation of micronaire with other non-cellulosic compounds suggesting that increased cellulose levels of high micronaire fibers affect the XG and HG epitopes in a different way than the other non-cellulosic epitopes. For instance, it could specifically decrease extractability of the XG and HG epitopes. As micronaire measures a combination of fiber fineness and maturity, we wanted to understand whether the observed correlation is with maturity or fineness or a combination of both. We tested this using linear regression models once again and built models for fineness and maturity of the fibers. We observed that the regression models for fineness had an adjusted  $R^2$  value of 0.803 with JIM5, LM19, and LM25 as significant predictors at a 1% threshold. The regression model for maturity was also significant at the 1% threshold but with no particular significant predictors thereby suggesting that the observed correlation is attributed to fiber fineness. This indicates that this correlation is linked to the thickness rather than the shape of the fiber, which is consistent with a link to the cellulose levels.

Since only the first canonical function of the CCA analysis is statistically significant and this function explains only for micronaire a large fraction of the variance, the results of the CCA analysis are not informative with respect to the other fiber properties. For these fiber properties, the correlation between fiber length and callose is the only one that was detected in both the linear regression and the sPLS analysis. Callose has been described to play a role in cotton fiber elongation. Indeed, it was reported that plasmodesmata closure was positively correlated with the rapid fiber elongation and that callose was involved in the gating of these plasmodesmata [76]. However, this observation involves transient callose detection, only after 5 dpa and already significantly reduced at 20 dpa, what makes it unlikely to be detected in mature fibers. Other callose deposition was reported by [77]. This callose is supposed to be deposited in the secondary cell wall and remains in the fiber. From the results of the multiple regression models (Table 2), a positive correlation between several of the homogalacturonan probes and length property of the fibers is apparent. The link between pectins and the elongation of cell walls is already observed in several plant systems [78] and studies in flax stems, pea stems and maize coleoptiles revealed a negative correlation between pectin levels and cell elongation. In cotton fibers and trichomes, there exists a positive correlation between pectic sheath and elongation [79] and recent studies by [80] have established that pectic polysaccharides and xyloglucan containing uronic acids were the major polysaccharides extracted during elongation. Hence, our results are in agreement with various studies which state that pectin biosynthesis promotes fiber elongation [81] and that the degree of esterification is a key factor in controlling the elongation [38,82]. The correlation between length and HG was not detected in the sPLS analysis most likely because the stronger (negative) correlation of HG with micronaire.

Furthermore, relationships between fiber strength or elongation and specific carbohydrate epitopes could be deduced from the results of the sPLS analysis (Figure 2). For instance, fiber strength was associated both with the xylan (LM11) and the extensin (JIM20) epitope. A role of xylan in fiber strength would be consistent with the function of heteroxylan in other cell types which is commonly related to the strengthening of cell walls as revealed by defects in cellulose deposition in xylan mutants [83]. A role of extensin in fiber strength is less expected and would need experimental validation. In the linear regression analysis, extensin was identified as a significant predictor for length AFIS but not for length HVI. A role for extensin in determining cotton fiber length would be more consistent with its role in other plant cell types [84]. Finally, AGP glycan (JIM13) and mannan (BS-400-4 and LM21) epitopes were found to predict cotton fiber elongation from the sPLS model. Interestingly, studies have indicated that AGPs are important players during fiber development. Immunofluorescence assays by JIM 13 showed distinct patterns in developing fiber cells indicating that polysaccharide chains of AGPs are involved in initiation and elongation stages of cotton fibers [85–87]. However, it is not clear how these AGPs would affect the elongation property of the mature fiber. These unexpected correlations present thus interesting hypotheses for further structure-function relationship studies of the cotton fiber.

Overall, CoMPP assays of cell wall polysaccharides from cotton fibers suggest that it will be a powerful tool in detecting and

quantifying the differences between large sets of cotton lines thus gathering lot of information which is necessary for a proper statistical approach. With the use of predictive statistical approaches to integrate different kinds of datasets, this analysis has thus discovered some correlations that are in line with already known biological functions and others for which the biological relevance still has to be tested. Also, it confirmed the relevance of this type of analysis to enable a detailed understanding of the data from CoMPP assays of cell wall polysaccharides. However, the use of mature cotton fibers in this analysis only allows detecting relevant correlations for components that are still present at maturity. In addition, many changes in polysaccharide composition occur between the fiber elongation stage and maturity. One would thus expect to identify only a fraction of the relationships between polysaccharide composition and fiber properties by analysis of mature fibers, especially for fiber properties such as length that are determined in the early stages of development. Hence it would be interesting to perform a similar kind of analysis using the polysaccharide composition of developing fibers to see whether additional relationships with fiber properties can be determined. The panel of cotton lines used in this study was selected to have maximal diversity in fiber properties and composition. Applying this type of analysis to commercially important cotton lines would allow to understand whether differences in polysaccharide composition affect properties of commercial cotton in the same way as observed in this study and to get insight into the developmental polysaccharides that are essential to obtain high quality cotton fibers. With the sequencing of the *G. hirsutum* genome, cotton fiber research is an exciting field and the work presented here will provide a base for future studies, with potential to translate this study on the developing fibers.

## Supporting Information

**Table S1 Fiber characteristics/phenotype measurements for the 32 cotton lines used in the study.** The plant introduction number (PI number) from the USDA national plant germplasm is also included for each cotton line. (XLSX)

**Table S2 Comprehensive microarray polymer profiling (CoMPP) heat map of CDTA and NaOH extractions of mature cotton fibers from 32 cotton lines.** References for probe specificity are listed in Table 1. (XLSX)

## Acknowledgments

We would like to thank Prof. JP Paul Knox, University of Leeds for the cell wall epitopes used in the glycan array experiment.

## Author Contributions

Conceived and designed the experiments: DR JLR XG WW FM JS. Performed the experiments: JLR XG. Analyzed the data: DR. Contributed reagents/materials/analysis tools: DR JLR XG. Contributed to the writing of the manuscript: DR JLR. Read and approved the manuscript: DR JLR XG WW FM JS.

## References

1. Heredia A, Jiménez A, Guillén R (1995) Composition of plant cell walls. *Z Für Lebensm-Unters -Forsch* 200: 24–31.
2. Keegstra K (2010) Plant Cell Walls. *Plant Physiol* 154: 483–486. doi:10.1104/pp.110.161240.
3. Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, et al. (2004) Toward a Systems Approach to Understanding Plant Cell Walls. *Science* 306: 2206–2211. doi:10.1126/science.1102765.

4. Minorsky PV (2002) The wall becomes surmountable. *Plant Physiol* 128: 345–353. doi:10.1104/pp.900022.
5. Carpita NC, Gibeault DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J Cell Mol Biol* 3: 1–30.
6. Roberts K (2001) How the Cell Wall Acquired a Cellular Context. *Plant Physiol* 125: 127–130. doi:10.1104/pp.125.1.127.
7. McCann M, Rose J (2010) Blueprints for Building Plant Cell Walls. *Plant Physiol* 153: 365–365. doi:10.1104/pp.110.900324.
8. Pilling E, Höfte H (2003) Feedback from the wall. *Curr Opin Plant Biol* 6: 611–616.
9. Somerville C (2006) Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol* 22: 53–78. doi:10.1146/annurev.cellbio.22.022206.160206.
10. Mutwil M, Debolt S, Persson S (2008) Cellulose synthesis: a complex complex. *Curr Opin Plant Biol* 11: 252–257. doi:10.1016/j.pbi.2008.03.007.
11. Ellis M, Egelund J, Schultz CJ, Bacic A (2010) Arabinogalactan-Proteins: Key Regulators at the Cell Surface? *Plant Physiol* 153: 403–419. doi:10.1104/pp.110.156000.
12. Chapple C, Carpita N (1998) Plant cell walls as targets for biotechnology. *Curr Opin Plant Biol* 1: 179–185. doi:10.1016/S1369-5266(98)80022-8.
13. Thakur BR, Singh RK, Handa AK (1997) Chemistry and uses of pectin—a review. *Crit Rev Food Sci Nutr* 37: 47–73. doi:10.1080/10408399709527767.
14. Sücklen MB (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nat Rev Genet* 9: 433–443. doi:10.1038/nrg2336.
15. Morris G, Kök S, Harding S, Adams G (2010) Polysaccharide drug delivery systems based on pectin and chitosan. *Biotechnol Genet Eng Rev* 27: 257–284.
16. Schena M (1996) Genome analysis with gene expression microarrays. *BioEssays News Rev Mol Cell Dev Biol* 18: 427–431. doi:10.1002/bies.950180513.
17. Ekins R, Chu FW (1999) Microarrays: their origins and applications. *Trends Biotechnol* 17: 217–218.
18. Wang D (2003) Carbohydrate microarrays. *Proteomics* 3: 2167–2175. doi:10.1002/pmic.200300601.
19. Park S, Lee M-R, Shin I (2008) Carbohydrate microarrays as powerful tools in studies of carbohydrate-mediated biological processes. *Chem Commun Camb Engl*: 4389–4399. doi:10.1039/b806699j.
20. Shin I, Park S, Lee M (2005) Carbohydrate microarrays: an advanced technology for functional studies of glycans. *Chem Weinh Bergstr Ger* 11: 2894–2901. doi:10.1002/chem.200401030.
21. Wang R, Liu S, Shah D, Wang D (2005) A practical protocol for carbohydrate microarrays. In: Zanders ED, editor. *Methods Mol Biol Clifton NJ* 310: 241–252.
22. Hsu T-L, Hanson SR, Kishikawa K, Wang S-K, Sawa M, et al. (2007) Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proc Natl Acad Sci U S A* 104: 2614–2619. doi:10.1073/pnas.0611307104.
23. Hanson SR, Hsu T-L, Weerapana E, Kishikawa K, Simon GM, et al. (2007) Tailored glycoproteomics and glycan site mapping using saccharide-selective bioorthogonal probes. *J Am Chem Soc* 129: 7266–7267. doi:10.1021/ja0724083.
24. Pedersen HL, Fangel JU, McCleary B, Ruzanski C, Rydahl MG, et al. (2012) Versatile high-resolution oligosaccharide microarrays for plant glycobiology and cell wall research. *J Biol Chem*: jbc.M112.396598. doi:10.1074/jbc.M112.396598.
25. Uchiyama N, Kuno A, Koseki-Kuno S, Ebe Y, Horio K, et al. (2006) Development of a lectin microarray based on an evanescent-field fluorescence principle. *Methods Enzymol* 415: 341–351. doi:10.1016/S0076-6879(06)15021-1.
26. Gupta G, Suroli A, Sampathkumar S-G (2010) Lectin microarrays for glycomic analysis. *Omic J Integr Biol* 14: 419–436. doi:10.1089/omi.2009.0150.
27. Gama CI, Tully SE, Sotogaku N, Clark PM, Rawat M, et al. (2006) Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* 2: 467–473. doi:10.1038/nchembio810.
28. De Paz JL, Noti C, Seeberger PH (2006) Microarrays of Synthetic Heparin Oligosaccharides. *J Am Chem Soc* 128: 2766–2767. doi:10.1021/ja057584v.
29. Ratner DM, Seeberger PH (2007) Carbohydrate microarrays as tools in HIV glycobiology. *Curr Pharm Des* 13: 173–183.
30. Wang L-X, Ni J, Singh S, Li H (2004) Binding of high-mannose-type oligosaccharides and synthetic oligomannose clusters to human antibody 2G12: implications for HIV-1 vaccine design. *Chem Biol* 11: 127–134. doi:10.1016/j.chembiol.2003.12.020.
31. Huang C-Y, Thayer DA, Chang AY, Best MD, Hoffmann J, et al. (2006) Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigen. *Proc Natl Acad Sci U S A* 103: 15–20. doi:10.1073/pnas.0509693102.
32. Lawrie CH, Marafioti T, Hatton CSR, Dirnhofer S, Roncador G, et al. (2006) Cancer-associated carbohydrate identification in Hodgkin's lymphoma by carbohydrate array profiling. *Int J Cancer J Int Cancer* 118: 3161–3166. doi:10.1002/ijc.21762.
33. Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, et al. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A* 101: 17033–17038. doi:10.1073/pnas.0407902101.
34. Liang P-H, Wang S-K, Wong C-H (2007) Quantitative analysis of carbohydrate-protein interactions using glycan microarrays: determination of surface and solution dissociation constants. *J Am Chem Soc* 129: 11177–11184. doi:10.1021/ja072931h.
35. Bryan MC, Wong C-H (2004) Aminoglycoside array for the high-throughput analysis of small molecule-RNA interactions. *Tetrahedron Lett* 45: 3639–3642. doi:10.1016/j.tetlet.2004.03.035.
36. Disney MD, Barrett OJ (2007) An aminoglycoside microarray platform for directly monitoring and studying antibiotic resistance. *Biochemistry (Mosc)* 46: 11223–11230. doi:10.1021/bi701071h.
37. Moller I, Marcus SE, Haeger A, Verherbruggen Y, Verhoef R, et al. (2008) High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles. *Glycoconj J* 25: 37–48. doi:10.1007/s10719-007-9059-7.
38. Singh B, Avci U, Eichler Inwood SE, Grimson MJ, Landgraf J, et al. (2009) A specialized outer layer of the primary cell wall joins elongating cotton fibers into tissue-like bundles. *Plant Physiol* 150: 684–699. doi:10.1104/pp.109.135459.
39. Aoki-Kinoshita KF, Kanehisa M (2006) Bioinformatics approaches in glycomics and drug discovery. *Curr Opin Mol Ther* 8: 514–520.
40. Von der Lieth C-W, Bohne-Lang A, Lohmann KK, Frank M (2004) Bioinformatics for glycomics: status, methods, requirements and perspectives. *Brief Bioinform* 5: 164–178.
41. Marchal I, Goffier G, Dugas O, Majed M (2003) Bioinformatics in glycobiology. *Biochimie* 85: 75–81.
42. Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM (2009) Evolution and Natural History of the Cotton Genus. In: Paterson AH, editor. *Genetics and Genomics of Cotton*. *Plant Genetics and Genomics: Crops and Models*. Springer US. 3–22.
43. Wang QQ, Liu F, Chen XS, Ma XJ, Zeng HQ, et al. (2010) Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/limitless mutant. *Genomics* 96: 369–376. doi:10.1016/j.ygeno.2010.08.009.
44. Al-Ghazi Y, Bourouf S, Arioli T, Dennis ES, Llewellyn DJ (2009) Transcript Profiling During Fiber Development Identifies Pathways in Secondary Metabolism and Cell Wall Structure That May Contribute to Cotton Fiber Quality. *Plant Cell Physiol* 50: 1364–1381. doi:10.1093/pcp/pcp084.
45. Gou J-Y, Wang L-J, Chen S-P, Hu W-L, Chen X-Y (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res* 17: 422–434. doi:10.1038/sj.cr.7310150.
46. Avci U, Pattathil S, Singh B, Brown VL, Hahn MG, et al. (2013) Cotton fiber cell walls of *Gossypium hirsutum* and *Gossypium barbadense* have differences related to loosely-bound xyloglucan. *PLoS One* 8: e56315. doi:10.1371/journal.pone.0056315.
47. Meikle PJ, Bonig I, Hoogenraad NJ, Clarke AE, Stone BA (1991) The location of (1→3)- $\beta$ -glucans in the walls of pollen tubes of *Nicotiana glauca* using a (1→3)- $\beta$ -glucan-specific monoclonal antibody. *Planta* 185: 1–8. doi:10.1007/BF00194507.
48. Willats WG, Limberg G, Buchholt HC, van Alebeek GJ, Benen J, et al. (2000) Analysis of pectic epitopes recognized by hybridoma and phage display monoclonal antibodies using defined oligosaccharides, polysaccharides, and enzymatic degradation. *Carbohydr Res* 327: 309–320.
49. Verherbruggen Y, Marcus SE, Haeger A, Ordaz-Ortiz JJ, Knox JP (2009) An extended set of monoclonal antibodies to pectic homogalacturonan. *Carbohydr Res* 344: 1858–1862. doi:10.1016/j.carres.2008.11.010.
50. Yates EA, Valdor JF, Haslam SM, Morris HR, Dell A, et al. (1996) Characterization of carbohydrate structural features recognized by anti-arabinogalactan-protein monoclonal antibodies. *Glycobiology* 6: 131–139.
51. Smallwood M, Beven A, Donovan N, Neill SJ, Peart J, et al. (1994) Localization of cell wall proteins in relation to the developmental anatomy of the carrot root apex. *Plant J* 5: 237–246. doi:10.1046/j.1365-3113.1994.05020237.x.
52. McCartney L, Marcus SE, Knox JP (2005) Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *J Histochem Cytochem Off J Histochem Soc* 53: 543–546. doi:10.1369/jhc.4B6578.2005.
53. Marcus SE, Verherbruggen Y, Hervé C, Ordaz-Ortiz JJ, Farkas V, et al. (2008) Pectic homogalacturonan masks abundant sets of xyloglucan epitopes in plant cell walls. *BMC Plant Biol* 8: 60. doi:10.1186/1471-2229-8-60.
54. Pettolino FA, Hoogenraad NJ, Ferguson C, Bacic A, Johnson E, et al. (2001) A (1→4)-beta-mannan-specific monoclonal antibody and its use in the immunocytochemical location of galactomannans. *Planta* 214: 235–242.
55. Marcus SE, Blake AW, Benians TAS, Lee KJD, Poyser C, et al. (2010) Restricted access of proteins to mannan polysaccharides in intact plant cell walls. *Plant J Cell Mol Biol* 64: 191–203. doi:10.1111/j.1365-3113.2010.04319.x.
56. R Core Team (2013) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
57. Tabachnick BG, Fidell LS (2012) *Using Multivariate Statistics*. 6 edition. Boston: Pearson. 1024 p.
58. Schneider A, Hommel G, Blettner M (2010) *Linear Regression Analysis*. *Dtsch Arztebl Int* 107: 776–782. doi:10.3238/arztebl.2010.0776.
59. Hair JF (2010) *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.
60. Lutz JG, Eckert TL (1994) The Relationship between Canonical Correlation Analysis and Multivariate Multiple Regression. *Educ Psychol Meas* 54: 666–675. doi:10.1177/0013164494054003009.
61. Tenenhaus A, Philippe C, Guillemot V, Cao K-AL, Grill J, et al. (2014) Variable selection for generalized canonical correlation analysis. *Biostatistics*. doi:10.1093/biostatistics/kxu001.

62. Witten DM, Tibshirani RJ (2009) Extensions of Sparse Canonical Correlation Analysis with Applications to Genomic Data. *Stat Appl Genet Mol Biol* 8.
63. Le Cao K-A, Martin PG, Robert-Granic C, Besse P (2009) Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* 10: 34. doi:10.1186/1471-2105-10-34.
64. Rencher AC (2003) Canonical Correlation. *Methods of Multivariate Analysis*. John Wiley & Sons, Inc. 361–379.
65. Thompson B (1984) Canonical correlation analysis uses and interpretation. Beverly Hills, Calif.: Sage Publications.
66. Dejean S, Gonzalez I, Lê Cao K-A with contributions from Monget P, Coquery J, Yao F, Lique B and Rohart F (2013) mixOmics: Omics Data Integration Project. R package version 5.0-1. Available: <http://CRAN.R-project.org/package=mixOmics>.
67. Boulesteix A-L, Strimmer K (2007) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform* 8: 32–44. doi:10.1093/bib/bbl016.
68. Lê Cao K-A, Rossouw D, Robert-Granic C, Besse P (2008) A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol* 7: Article 35. doi:10.2202/1544-6115.1390.
69. González I, Déjean S, Martin PGP, Gonçalves O, Besse P, et al. (2009) Highlighting relationships between heterogeneous biological data through graphical displays based on regularized canonical correlation analysis. *J Biol Syst* 17: 173–199. doi:10.1142/S0218339009002831.
70. González I, Cao K-AL, Davis MJ, Déjean S (2012) Visualising associations between paired “omics” data sets. *BioData Min* 5: 19. doi:10.1186/1756-0381-5-19.
71. Chun H, Keles S (2010) Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Ser B Stat Methodol* 72: 3–25. doi:10.1111/j.1467-9868.2009.00723.x.
72. Bowman MJ, Park W, Bauer PJ, Udall JA, Page JT, et al. (2013) RNA-Seq Transcriptome Profiling of Upland Cotton (*Gossypium hirsutum* L.) Root Tissue under Water-Deficit Stress. *PLoS One* 8: e82634. doi:10.1371/journal.pone.0082634.
73. Lacape J-M, Claverie M, Vidal RO, Carazzolle MF, Guimarães Pereira GA, et al. (2012) Deep Sequencing Reveals Differences in the Transcriptional Landscapes of Fibers from Two Cultivated Species of Cotton. *PLoS ONE* 7: e48855. doi:10.1371/journal.pone.0048855.
74. Rambani A, Page JT, Udall JA (2014) Polyploidy and the petal transcriptome of *Gossypium*. *BMC Plant Biol* 14: 1–14. doi:10.1186/1471-2229-14-3.
75. Gilbert MK, Turley RB, Kim HJ, Li P, Thyssen G, et al. (2013) Transcript profiling by microarray and marker analysis of the short cotton (*Gossypium hirsutum* L.) fiber mutant Ligon lintless-1 (Li 1). *BMC Genomics* 14: 403. doi:10.1186/1471-2164-14-403.
76. Ruan Y-L, Xu S-M, White R, Furbank RT (2004) Genotypic and Developmental Evidence for the Role of Plasmodesmatal Regulation in Cotton Fiber Elongation Mediated by Callose Turnover. *Plant Physiol* 136: 4104–4113. doi:10.1104/pp.104.051540.
77. Sahnikov VV, Grimson MJ, Seagull RW, Haigler CH (2003) Localization of sucrose synthase and callose in freeze-substituted secondary-wall-stage cotton fibers. *Protoplasma* 221: 175–184. doi:10.1007/s00709-002-0079-7.
78. Goldberg R, Morvan C, Jauneau A, Jarvis MC (1996) Methyl-esterification, de-esterification and gelation of pectins in the primary cell wall. In: J Visser and A.G.J Voragen, editor. *Progress in Biotechnology. Pectins and Pectinases Proceedings of an International Symposium*. Elsevier, Vol. Volume 14. 151–172.
79. Vaughn KC, Turley RB (1999) The primary walls of cotton fibers contain an ensheathing pectin layer. *Protoplasma* 209: 226–237. doi:10.1007/BF01453451.
80. Tokumoto H, Wakabayashi K, Kamisaka S, Hoson T (2002) Changes in the sugar composition and molecular mass distribution of matrix polysaccharides during cotton fiber development. *Plant Cell Physiol* 43: 411–418.
81. Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci* 3. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3356883/>. Accessed 24 Jun 2014.
82. Wang H, Guo Y, Lv F, Zhu H, Wu S, et al. (2010) The essential role of GhPEL gene, encoding a pectate lyase, in cell wall loosening by depolymerization of the de-esterified pectin during fiber elongation in cotton. *Plant Mol Biol* 72: 397–406. doi:10.1007/s11103-009-9578-7.
83. Hao Z, Mohnen D (2014) A review of xylan and lignin biosynthesis: foundation for studying Arabidopsis irregular xylem mutants with pleiotropic phenotypes. *Crit Rev Biochem Mol Biol* 49: 212–241. doi:10.3109/10409238.2014.889651.
84. Sadava D, Chrispeels MJ (1973) Hydroxyproline-rich cell wall protein (extensin): Role in the cessation of elongation in excised pea epicotyls. *Dev Biol* 30: 49–55. doi:10.1016/0012-1606(73)90047-X.
85. Huang G-Q, Gong S-Y, Xu W-L, Li W, Li P, et al. (2013) A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. *Plant Physiol* 161: 1278–1290. doi:10.1104/pp.112.203760.
86. Qin L-X, Rao Y, Li L, Huang J-F, Xu W-L, et al. (2013) Cotton GalT1 Encoding a Putative Glycosyltransferase Is Involved in Regulation of Cell Wall Pectin Biosynthesis during Plant Development. *PLoS ONE* 8: e59115. doi:10.1371/journal.pone.0059115.
87. Bowling AJ, Vaughn KC, Turley RB (2011) Polysaccharide and glycoprotein distribution in the epidermis of cotton ovules during early fiber initiation and growth. *Protoplasma* 248: 579–590. doi:10.1007/s00709-010-0212-y.