รายงานวิจัยฉบับสมบูรณ์

ชีววิทยาระบบของการแสดงออกทางพันธุกรรมในพืช เพื่อตอบสนองต่อการเปลี่ยนแปลงอุณหภูมิ

ดร.วโรดม เจริญสวรรค์ และคณะ

พฤษภาคม 2560

สัญญาเลขที่ TRG5880067

รายงานวิจัยฉบับสมบูรณ์

ชีววิทยาระบบของการแสดงออกทางพันธุกรรมในพืช เพื่อตอบสนองต่อการเปลี่ยนแปลงอุณหภูมิ

ดร.วโรดม เจริญสวรรค์
คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
และมหาวิทยาลัยมหิดล

Abstract

Project Code : TRG5880067

Project Title : A system-wide analysis of transcriptional regulation in hybrid plants in response to temperature change

Investigator : Dr. Varodom Charoensawan *et al.* Faculty of Science, Mahidol University

E-mail Address : varodom.cha@mahidol.ac.th

Project Period : 2 years

Keywords : Systems Biology, Regulation of Gene Expression, Climate Change, Plant Molecular Biology, High-throughput Sequencing

Transcription in eukaryotes is tightly regulated by the interplay between the proteins transcription factors (TFs) and nucleosomal histones. It has been demonstrated that the eukaryotic TF Heat Shock Factor 1 (HSF1), and the histone variant H2A.Z both play a major role in mediating transcription in response to temperature changes, one of the most important external stimuli, especially in the light of the extreme environment due to climate changes. It is not clear, however, how these two proteins interplay in this important transcriptional regulation process. We investigate this long-standing question using a model plant Arabidopsis thaliana (wild-type plants and hybrids). Plants have to adapt to fluctuating temperatures both diurnally and seasonally, and thus serve as a useful model for this particular genome-environment interaction question. We have generated a large-scale dataset of transcriptomes (RNA-seq) and H2A.Z and HSF1 occupancy profiles (ChIP-seq) of wild-type plants shifted to different ambient temperatures, in order to explore how the local and global changes in occupancies of H2A.Z-containing nucleosome and the HSF1 TF, affect transcriptional readouts. To investigate the link between the genomic interaction in hybrids and temperature changes, we also study the Arabidopsis accessions Col-*0*, C24 and hybrids grown low (22^{o}C) and high (27^{o}C) ambient temperatures. The differential growth rates of the hybrids are calculated and linked with corresponding time-course transcriptomic profiles, in order to identify a set of genes that control differential growth rate in hybrids and their responsiveness to temperature changes.

บทคัดย่อ

รหัสโครงการ : TRG5880067

ชื่อโครงการ : ชีววิทยาระบบของการแสดงออกทางพันธุกรรมในพืชเพื่อตอบสนองต่อการเปลี่ยนแปลงอุณหภูมิ

ชื่อนักวิจัย และสถาบัน ดร.วโรดม เจริญสวรรค์ และคณะ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล

E-mail Address : varodom.cha@mahidol.ac.th

ระยะเวลาโครงการ: 2 ปี

คำหลัก : ชีววิทยาระบบ, การควบคุมการแสดงออกทางพันธุกรรม, สภาวะอากาศเปลี่ยนแปลง, ชีวโมเลกุลในพืช, การหาลำดับเบสปริมาณมาก

ทรานสปริปชัน (Transcription) เป็นขั้นตอนสำคัญในการแสดงออกทางพันธุกรรมของสิ่งมีชีวิตทุกชนิด ซึ่งถูกควบคุมด้วยโปรตีนหลายชนิดเช่น ทรานสปริปชัน แฟคเตอร์ (transcription factors, TFs) และ ฮิสโตน (histones) ในการควบความการแสดงออกทางพันธุกรรมเพื่อตอบสนองต่อการเปลี่ยนแปลงอุณหภูมิ สิ่งมีชีวิตอาศัย TF Heat Shock Factor 1 (HSF1) และ histone variant H2A.Z ทำงานร่วมกัน แต่กลไลในการทำงานของทั้งสองโปรตีนนั้นไม่ชัดเจน ในงานวิจัยชิ้นนี้ผู้วิจัยได้ศึกษาปัญหาดังกล่าวในพืช Arabidopsis thaliana (ในพ่อแม่พันธุ์และในลูกผสม) เนื่องจากสภาวะอากาศที่เปลี่ยนแปลงมีผลกระทบอย่างมากต่อผลผลิตทางเกษตร ผู้วิจัยใช้เทคโนโลยีการศึกษาข้อมูลทางชีวโมเลกุลขนาดใหญ่ RNA-seq และ ChIP-seq และได้แสดงให้เห็นถึงการปฏิสัมพันธ์ในลักษณะต่าง ๆของโปรตีนทั้งสองชนิด และผลของการควบคุมการแสดงออกในรูปแบบของระดับ transcription เมื่ออุณหภูมิเพิ่มขึ้น และยังศึกษาปัญหาดีงกล่าวในลูกผสม ซึ่งเป็นหนึ่งในวิธีที่ใช้มากในการเพิ่มพืชผลทางเกษตร เพื่อให้เกิดความเข้าใจว่าพืชทั่วไปและพืชลูกผสมได้รับผลกระทบอย่างไรต่อการเปลี่ยนแปลงสภาวะอากาศ เพื่อให้ใช้การพัฒนาสายพันธุ์ให้เหมาะสมต่อไป

A system-wide analysis of transcriptional regulation in hybrid plants in response to temperature change

Objectives

1. To establish a system-wide transcriptional profile of plants subjected to ambient temperature changes, which will serve as a platform for investigating global and local regulatory mechanisms of temperature transcriptomes.
2. To elucidate the interplay between temperature-responsive phenotype and genetic effect using hybrid plants as a model, and identify differentially expressed genes between hybrids and parents.
3. To design and construct an analytic pipeline for next-generation sequencing (NGS) data from publicly available computer codes and applications (e.g. for RNA-seq, ChIP-seq), which can also be adapted and implemented on other genomes in the future.
4. To initiate collaborative research projects between systems biologists and experimental researchers, using high-throughput experimental and analytical methods as a common ground.

Methodology and Materials

As proposed, we have adopted a Systems Biology approach to globally characterize the genome-wide transcriptional patterns of wild-type plants and hybrids in response to temperature changes. This interdisciplinary problem-based project will combine several experimental as well as computational tools and techniques to investigate the transcriptional profiles and regulatory mechanisms, including Computational Biology, Bioinformatics, Molecular Biology, Plant Biology and Genetics, and Transcriptomics. We have also extended the pipeline initially proposed for transcriptomic analyses in the model plant *Arabidopsis thaliana,* to other model species including human and budding yeast. This also resulted in collaborative projects with experimental scientists and clinicians, as described below.

1. Plant materials and growth conditions

The *Arabidopsis thaliana* accessions Col-0 and C24 were crossed to produce F1 hybrid seeds. The seeds of F1 hybrids, parents, and *pif4-101* (a mutant lacking functional PIF4) were grown at 17^{o}C, 22^{o}C, and 27^{o}C in the short-day photoperiod, where the hypocotyl growth and PIF4 activity is maximized (Kumar 2010; Wigge 2013). Light intensity is controlled to minimize potential crosstalk effect of light and temperature sensing pathways. Five-day-old seedlings were imaged using a high-regulation scanner, and the hypocotyl lengths will be measured using Image. Statistical analysis of hypocotyl growth were performed in the statistical packages R (R core team, http://www.R-project.org). The experimental results of plant growth were intitially obtained at the Sainsbury Laboratory Cambridge University, UK under advice of Dr. Philip A Wigge, in collaboration with Drs. Sandra Cortoji and Matt Box. The computational analyses were implemented and performed by Ms. Napaporn Sriden, a Ph.D. student at Department of Biochemistry, Faculty of Science, Mahidol University, under the supervision of Dr. Varodom Charoensawan.

2. Construction of temperature transcriptomes

Whole seedlings were collected and used for RNA extraction. RNA quality were accessed using the BioAnalyzer or TapeStation systems (Agilent), as high quality RNA with minimal degradation is recommended for subsequent sequencing library preparation. The RNA-seq library were performed using the Illumina library preparation kits (*e.g.* TruSeq), to ensure compatibility with NGS technology that will be used at later stage. In brief, mRNA with polyA tails were selectively extracted, fragmented, and reverse transcribed into cDNA, which were then primed with sequencing adaptors and indices. The prepared sequencing libraries were analyzed using 100 base pair pair-end format (100 bps are sequenced from both directions of reads), using the Illumina HiSeq2000. The NGS library preparation were performed conducted at the Sainsbury Laboratory Cambridge University, UK, which were supported by the European Research Council (ERC) through the Wigge laboratory. The transcriptome-generating pipeline was re-implemented at Mahidol University, Thailand, using the funding from the Thailand Research Fund (TRF) by Dr. Varodom Charoensawan. The transcriptomic analyses were implemented and performed by Ms. Napaporn Sriden, a Ph.D. student at Department of Biochemistry, Faculty of Science, Mahidol University, under the supervision of Dr. Varodom Charoensawan, as described in more details below.

3. Computational analyses of Next-Generation Sequencing (NGS) data, and bioinformatics analyses

We have successfully developed an in-house NGS analytic pipeline, by combining a number of computational codes and applications publicly available. The pipeline for RNA-seq library has been robustly tested and used in several on-going projects at the Sainsbury Laboratory Cambridge University, and have now been re-implemented at the Integrative Computational BioScience (ICBS) center, Mahidol University.

In summary, raw reads obtained from the NGS sequencing service go through the quality control process to check for adaptor sequences, optical duplication rates, GC contents, using in-house Linux shell scripts that format and manage large-scale data and run FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We use Trimmomatic and/or FASTX (http://hannonlab.cshl.edu/fastx_toolkit/) to trim adaptor sequences that may appear in the reads, if the insert size is shorter than the sequenced length. Trimmed reads are then aligned to reference genome using Bowtie (Langmead 2012) and Tophat (Trapnell 2009). The normalization and estimation of relative transcriptional abundance are performed using Cufflinks and Cuffdiffs (Trapnell 2010). Mapped reads can also be visualized and analyzed using a genome browser, such as the integrative genome viewer (IGV) (Thorvaldsdottir 2013). As explained, these subsequent bioinformatic and statistical analyses have been performed in-house at Mahidol University, Thailand, using purposely written scripts, implemented in R, Perl, and Python (see Figure 1).

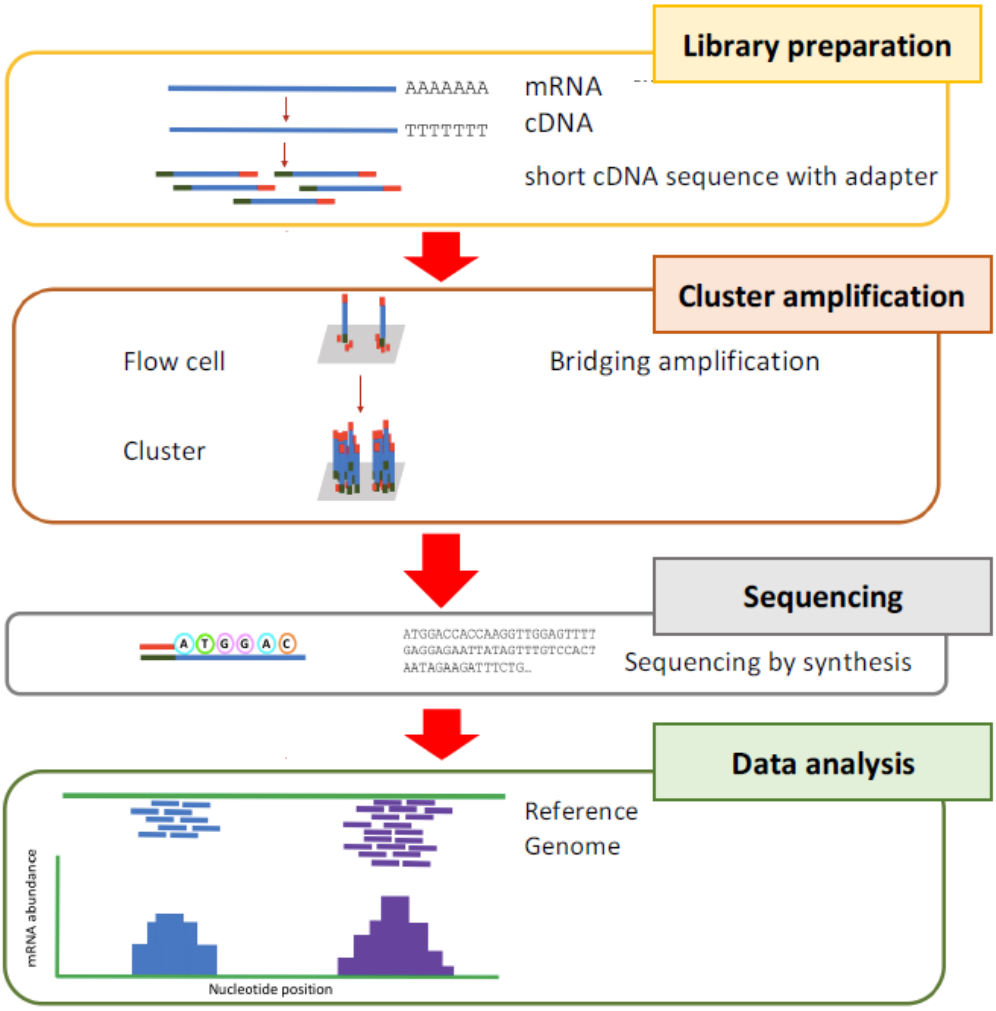


Figure 1: Summary of RNA-seq analytic pipelines. The illustration was created by Ms. Napaporn Sriden and was adapted from Wang and coworkers (Wang 2009).

4. Implementation of systems biology and network biology tools

One of the main objectives of this project is to establish analytical pipelines for high-throughput biological big data, as well as a common ground between systems biology, and experimental scientists and clinicians. In doing so, we used network analytic tool Cytoscape (http://www.cytoscape.org/) to showcase how large-scale biological data can be represented and analyzed. As a result, we have formed successfully collaborations with the laboratories of Prof. Dr. Sawawut Jitrapakee and Dr. Natini Jinawath, to work on network biology of miRNA in metabolic regulation and cancer biology.

Results and Discussion

RNA-seq analytic pipeline has successfully implemented

As proposed, we have successfully re-implemented the in-house NGS analytical pipelines (including RNA-seq) on the high-throughput computing (HPC) facility at the Integrative Computational BioScience (ICBS) center, Mahidol University. The pipelines consist of a set of publicly available libraries (as described in Methodology), were created and implemented by Dr. Varodom Charoensawan.

An example of mapped reads is shown in Figure 2, which describes higher numbers of reads mapped to the Heat Shock Protein 70 (HSP70) locus at higher temperatures (Charoensawan, Cortijo, and Wigge, under review).

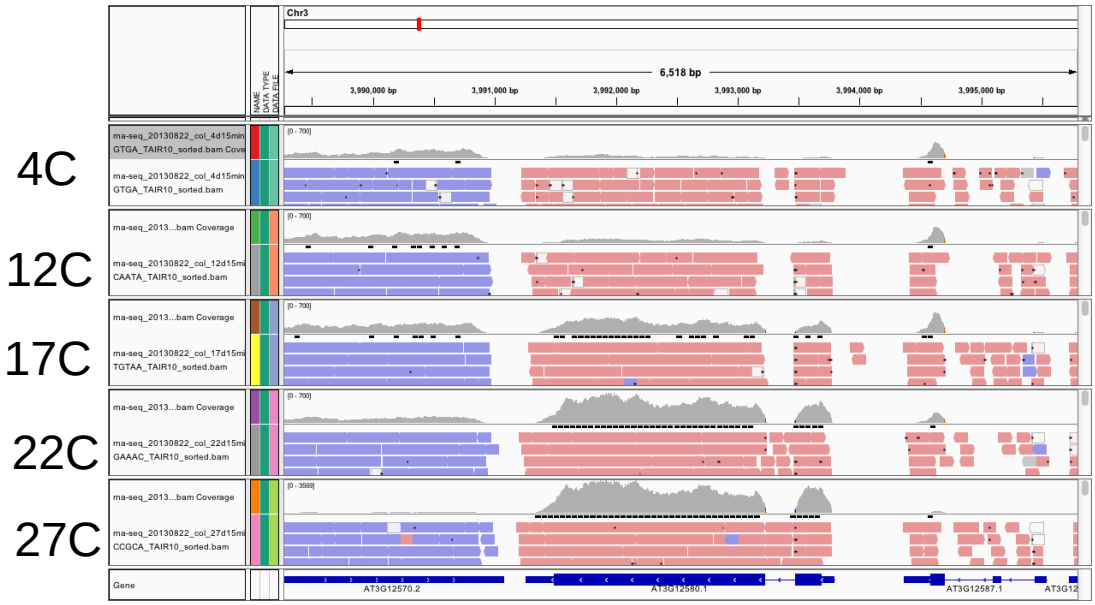


Figure 2: Snapshot from IGV describing upregulation of HSP70 transcript at higher ambient temperature. Blue indicates reads mapped to the forward strand, and red indicates reads mapped to the reverse strand.

The warm temperature transcriptome is highly dynamic

Using our time-course RNA-seq of Arabidopsis seedings grown at 17°C and 27°C, we have observed specific set of genes with highly temperature responsiveness (Cluster 6 in Figure 3 below). These genes are enriched in stress and environmental responses, including HSP70, a heat shock protein known to be induced by high ambient temperature (Kumar and Wigge 2010). All transcriptomic levels were estimated by normalization with the length of genes and number of reads in each sample (e.g. TPM, transcripts per million reads). Red cells depict up-regulated and green for down-regulated as compared to the zero time-point (before shifting from 17oC and 27oC).

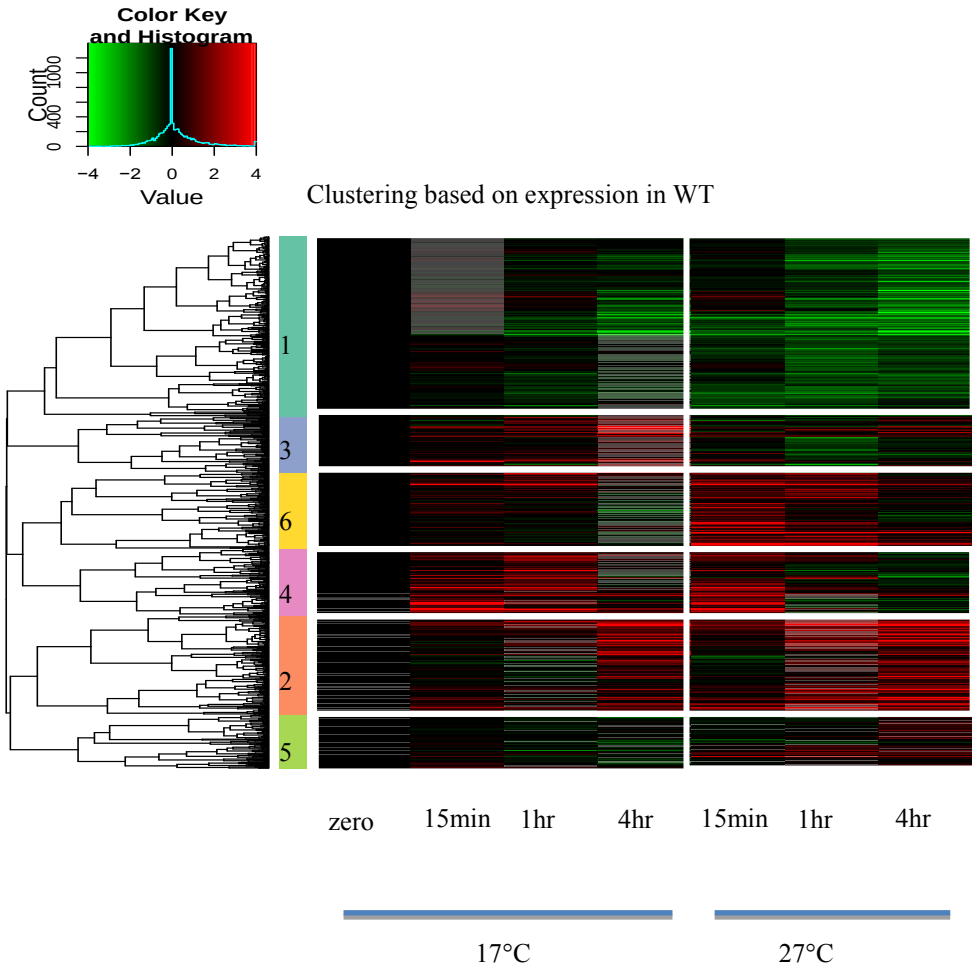


Figure 3: Heatmap demonstrating relative changes of transcriptional level of temperature responsive genes. Red and green indicate up- and down-regulated genes, as compared to time-point zero.

Temperature transcriptional transcriptional induction in plants is associated with H2A.Z eviction

The expression of heat responsive genes such as *HSP70* is accompanied by the eviction of nucleosomes containing the alternative histone H2A.Z, but the mechanism by which this occurs is not known. We performed micrococcal nuclease (MNase) digestion followed by high-throughput sequencing on seedlings subjected to identical temperature shifts as described above. Nucleosomes protect the DNA from MNase cleavage and hence MNase accessibility was used to infer nucleosome occupancy. In addition, we also performed ChIP of HTA11-tagged to locate the binding position and occupancy of the HTA11 protein, one of the three H2A.Z proteins in plants. We observe a noticeable reduction of ChIP signal in the plants shifted from 17°C to 27°C for 15 min among the Cluster 6 genes, consistent with this region becoming accessible to the transcriptional machinery upon increased temperature. There is a clear link between the loss of H2A.Z-nucleosomes and MNase-seq signals around +1 nucleosome position in the genes in Cluster 6 (see Figure 4 below), when

shifted from 17°C to 27°C, suggesting the genomic DNA becomes more accessible when H2A.Z (represented by HTA11 ChIP signal) is evicted.

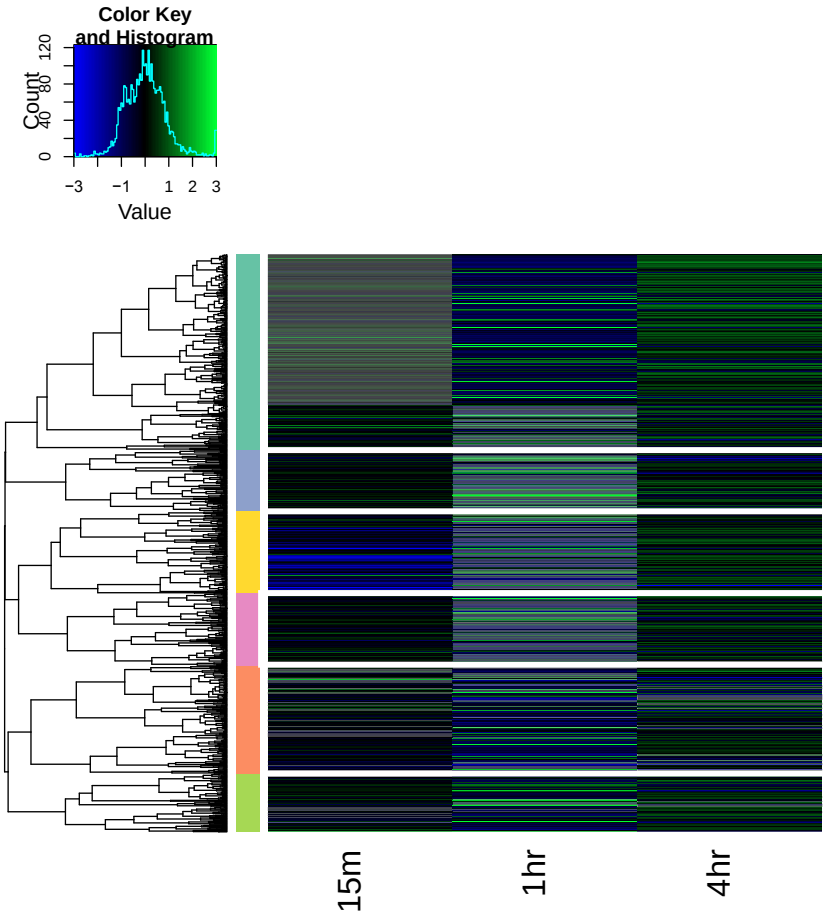


Figure 4 Heatmap demonstrating relative changes of HTA11 ChIP signal at the temperature responsive genes. Blue and green indicate increases and decrease in ChIP signal at 27°C, as compared to 17°C.

Phenotype of seedling growth at different temperatures

In addition to transcriptomic changes, we also investigated phenotypic variations of plants grown at different temperature. On top of the wild-type plant previously used before, Col-*0*, we also introduced to the study, different accession plant C24, and their hybrids, also grown at different temperatures, in order to observe the relationship between genetic and environmental factors on their growth. This was performed by taking photos of plants under infrared, and image analysis of the plant growth in three days (as in examples below). The plants were growth at 22°C and 27°C. This work was performed at the Sainsbury Laboratory Cambridge University, in collaboration with Drs. Sandra Cortoji and Matt Box.

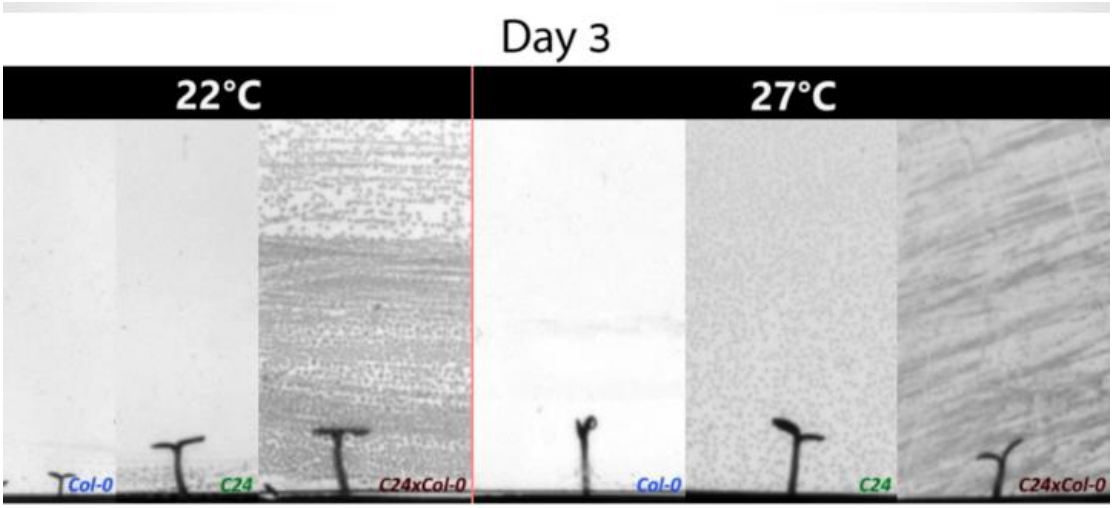


Figure 5: Infrared (IR) images of plants of different accession and hybrid grown at different ambient temperatures.

Growth rate of seedlings is maximized just before dawn

In line with what previously observed in (Box 2015), we have observed that seedlings grow the fastest just before dawn, when the growth quickly drops due to inactivation of the functions of growth related proteins such as PIF4 (growth rate of Col-*0* shown in Figure 6). Remarkably, we have seen that the hybrid C24xCol-*0* demonstrate the most steep growth also at the time point ZT0, or just before dawn. This suggests that this time point would be particularly interesting to investigate for transcriptional level of genes differentially transcribed in parents and in hybrids, which might be related to the regulation of heterosis in these lines (Figure 7).

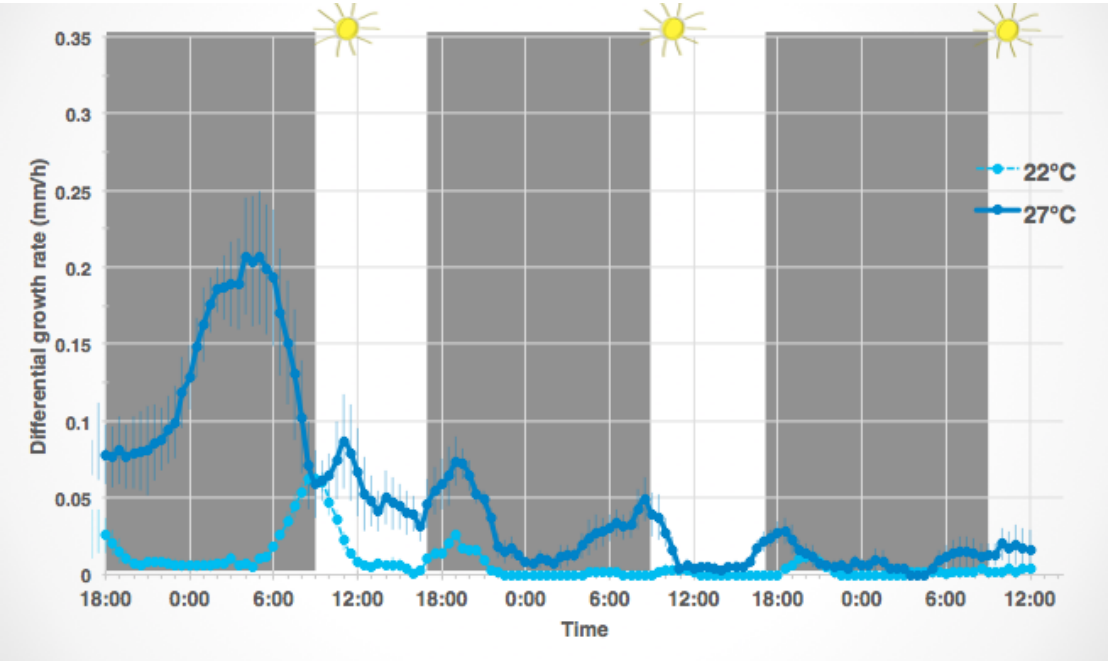


Figure 6: Differential growth rate of Col-0 grown at 22°C and 27°C for three days. The rate was calculated from IR images taken from 2 days after sowing.

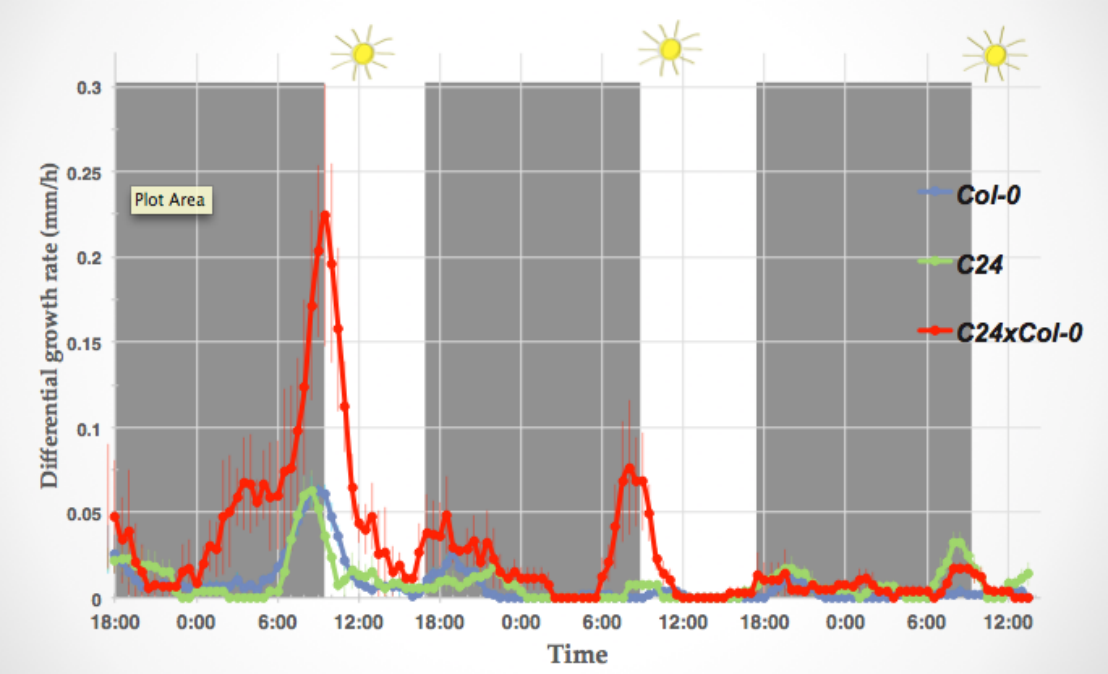


Figure 7: Differential growth rate of Col-0, C24, and hybrid grown at 22°C and 27°C for three days. The rate was calculated from IR images taken from 2 days after sowing.

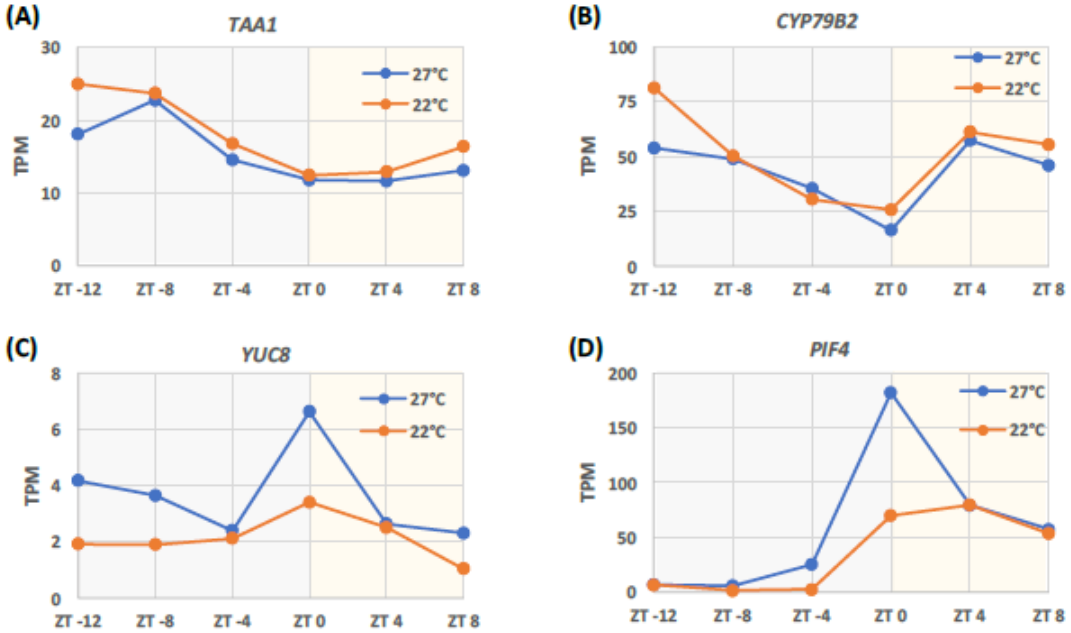


Figure 8: Time-course differential expression in Col-0 during ZT-12 – ZT8 of 5 days old seedlings. Grey area indicates the night time, whereas yellow area indicates day time. (A)TAA1, (B) CYP79B2, (C) YUC8 and (D) PIF4.

The differential expression in the key important gene in response to high ambient temperature

In addition to the phenotypic observation of plants under an elevated ambient temperature (27°C), the time-course transcriptomes were also performed to investigate the molecular basis of this phenomenon. Time-course RNA-seq libraries were prepared by Drs. Varodom Charoensawan, Matt Box, Sandra Cortijo at Sainsbury Laboratory as described in Methodology. The sequenced NGS reads were analyzed by Ms. Napaporn Sriden under Dr. Charoensawan's supervision. The time-course transcriptomic analysis reveals different dynamic patterns of transcription among selected genes. Previous studies from (Koini 2009; Leibman 2014) showed that auxin biosynthesis genes, YUC8, TAA1 and CYP79B2, were up-regulated in Arabidopsis thaliana growing under a high ambient temperature. Here, the time-course transcriptome of Col-*0* from this study showed two distinct transcription dynamic patterns (Figure 8). We found that high ambient temperature promoted only the transcription of YUC8 at all time points, without significantly changing the dynamic patterns of transcriptional changes. In contrast, transcription level of TAA1 and CYP79B2 at 22°C was slightly greater than the transcription level at 27°C. Transcriptional level of TAA1 appeared to decrease at nighttime until dawn, then slightly increase again during the day; whereas CYP79B2's transcription gradually dropped until dawn, then suddenly peaked up 4 hours after light. For YUC8, its transcription dynamics showed a similar pattern in comparison to the differential growth of Col-*0*, which gradually increased at night and the peak appeared just before dawn.

Conclusion

We have successfully implemented a number of NGS analytic pipelines for investigating transcriptomic changes (RNA-seq) and genome-wide DNA-protein interaction (ChIP-seq) at Mahidol University, Thailand. This analytical competency is not only a key to systems biology work of the Charoensawan laboratory, as initially proposed in the project, but also provides an opportunity for the group to collaborate with established research groups, in order to combine multidisciplinary expertise to investigate existing biological problems using a new approach. Figure 9 demonstrates a network of multiple miRNAs and oncogenic transcription factors controlling metabolic reprogramming in cancers, which normally would be studied individually without the knowledge and tools from systems biology.

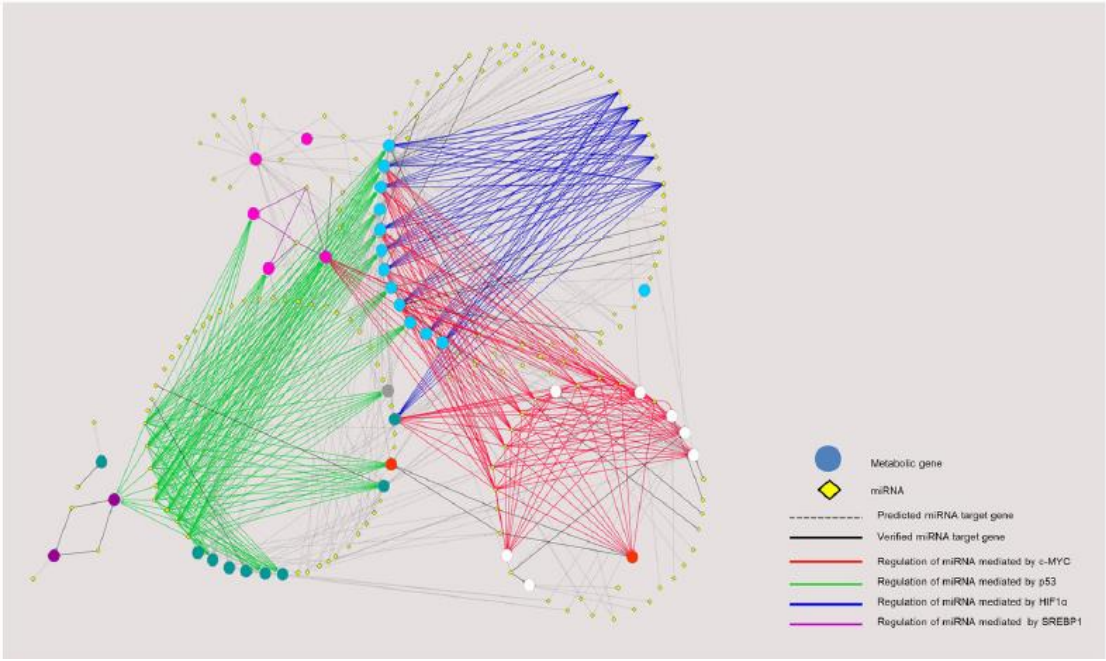


Figure 9: Regulatory network of miRNAs and oncogenic transcription factors controlling metabolic reprogramming in cancers. The figure shows direct and indirect miRNAs-metabolic genes interaction. The miRNAs that have already verified their regulatory function show in solid edges whereas the dash edges represent the overlap miRNAs from predictions only. The figure depicts a network representation, an important tool of Systems Biology that can be used to represent multiple biomolecular entities. The figure was adapted from a review by Pinweha and colleagues (Pinweha 2016), where Dr. Charoensawan is a co-author.

Molecular mechanisms of temperature sensing and response in plants and other species are crucial to mitigation of global climate changes. However, we are far from thorough understanding for this biological process, and thus adapting to this big change. Focusing on our transcriptomic results of RNA-seq from plants grown under different temperature, we are able to globally investigate a set of temperature-responsive genes that are highly inducible by the interplay of H2A.Z and HSF1. In

collaboration with Drs. Philip A Wigge and Sandra Cortigo, we have submitted a manuscript (where Dr. Charoensawan is a co-first author and co-corresponding author), which currently being revised. Additional transcriptomic investigation of plants have already been performed and being analyzed by Ms. Napaporn Sriden, a Ph.D. student under Dr. Charoensawan's supervision. So far, we have observed diurnal growth of Arabidopsis seedlings using the IR imaging described in the results, and are now moving to focus on the time-course transcriptomic data of wide-type parents and hybrids.

Future Directions

We have successfully established a pipeline and facility for analysis of high-throughput biological data (RNA-seq and ChIP-seq), and have started to utilize the competency to help and collaborate with other research groups to analyze their **"big data".** Moving on to the era of **"Thailand 4.0",** we foresee that biomolecular and clinical research will move from investigation individual genes, proteins, or diseases, to **automatically identification of biomarkers, biological pathways** that are important to certain biological problems **using large-scale biological data**. We are confident that the core competency established here will play a key role in assisting our research community toward this new direction.

Focusing on our molecular basis of temperature-response transcriptomes in plants and hybrids, we have now separately analyzed the plants' phenotypic and transcriptomic changes, and will now combine the two types of data to form a comprehensive conceptual model that can explain how hybrid vigors are achieved and how these are affected by temperature changes.

References

Box MS, Huang BE, Domijan M, Jaeger KE, Khattak AK, Yoo SJ, Sedivy EL, Jones DM, Hearn TJ, Webb AA, Grant A, Locke JC, Wigge PA. ELF3 controls thermoresponsive growth in Arabidopsis. *Curr Biol.* 2015 19;25(2):194-9.

Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nature methods* 2012, 9(4):357-359.

Leibman M, Shryock JJ, Clements MJ, Hall MA, Loida PJ, McClerren AL, et al. Comparative analysis of maize (Zea mays) crop performance: natural variation, incremental improvements and economic impacts. *Plant Biotechnology Journal.* 2014;12(7):941-50.

Koini MA, Alvey L, Allen T, Tilley CA, Harberd NP, Whitelam GC, et al. High Temperature-Mediated Adaptations in Plant Architecture Require the bHLH Transcription Factor PIF4. *Current Biology*. 2009. 1 9(5):408-13.

Kumar SV, and Wigge PA. H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell.* 2010;140, 136-147.

Thorvaldsdottir H, Robinson JT, Mesirov JP: Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics 2013, 14(2):178-192.

Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009, 25(9):1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 2010, 28(5):511-515.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.

Wigge PA: Ambient temperature signalling in plants. *Cur opin plant biol* 2013, 16(5):661-666.

Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (ระบุชื่อผู้แต่ง ชื่อเรื่อง ชื่อวารสาร ปี เล่มที่ เลขที่ และหน้า) หรือผลงานตามที่คาดไว้ในสัญญาโครงการ

1. Pinweha P, Rattanapornsompong K, **Charoensawan V**, Jitrapakdee S. MicroRNAs and Oncogenic Transcriptional Regulatory Networks Controlling Metabolic Reprogramming in Cancers. *Computational and Structural Biotechnology Journal.* 2016. Supported by TRG5880067 (SCOPUS SJR = 0.752, Q2).

2. Yang W, Schuster C, Beahan CT, **Charoensawan V**, Peaucelle A, Bacic A, Doblin MS, Wightman R, Meyerowitz EM. Regulation of Meristem Morphogenesis by Cell Wall Synthases in Arabidopsis. *Current Biology.* 2016. Supported by TRG5880067. (IF = 9.571, Q1).

3. Tanramluk D, Narupiyakul L, Akavipat R, Gong S, **Charoensawan V**. MANORAA (Mapping Analogous Nuclei Onto Residue And Affinity) for identifying protein-ligand fragment interaction, pathways and SNPs. *Nucleic Acids Research*. (2016). Supported by TRG5880067. (IF = 9.112, Q1).

4. Jinawath N, Bunbanjerdsuk S, Chayanupatkul M, Ngamphaiboon N, Asavapanumas N, Svasti J, **Charoensawan V*.** Bridging the gap between clinicians and systems biologists: from network biology to translational biomedical research. *Journal of Translational Medicine.* (2016) Supported by TRG5880067. (IF = 3.694, Q1).

In addition, we have one manuscript submitted and under revision.

1. Cortijo S*, **Charoensawan V***, Brestovitsky A, Buning R, Ravarani C, Rhodes D, van Noort J, Jaeger KE, Wigge PA. The daytime warm temperature transcriptome is controlled by the interplay of H2A.Z-nucleosomes and HSF transcription factors in Arabidopsis *(under revision).*

2. การนำผลงานวิจัยไปใช้ประโยชน์
   - เชิงพาณิชย์ (มีการนำไปผลิต/ขาย/ก่อให้เกิดรายได้ หรือมีการนำไปประยุกต์ใช้โดยภาคธุรกิจ/บุคคลทั่วไป)
   - เชิงนโยบาย (มีการกำหนดนโยบายอิงงานวิจัย/เกิดมาตรการใหม่/เปลี่ยนแปลงระเบียบข้อบังคับหรือวิธีทำงาน)
   - เชิงสาธารณะ (มีเครือข่ายความร่วมมือ/สร้างกระแสความสนใจในวงกว้าง)
   - เชิงวิชาการ (มีการพัฒนาการเรียนการสอน/สร้างนักวิจัยใหม่)

Please see attached document

3. อื่นๆ (เช่น ผลงานตีพิมพ์ในวารสารวิชาการในประเทศ การเสนอผลงานในที่ประชุมวิชาการ หนังสือ การจดสิทธิบัตร)

Invited Talk

1. Charoensawan V. Exploring plant's temperature transcriptome and its epigenetic regulation. Genomics and Systems Biology VI, New York University at Abu Dhabi, UAE. February 2016.
2. Charoensawan V. Exploring Temperature Transcriptional Regulation: A Systems Biology Approach. 5th Biochemistry and Molecular Biology conference, Songkhla, Thailand. May 2016.

Poster presentation

1. Napaporn Sriden, Sandra Cortijo, Matt Box, Philip A Wigge, Varodom Charoensawan. Regulation of Temperature Transcriptomes in Hybrid Plants. 5th Biochemistry and Molecular Biology conference, Songkhla, Thailand. May 2016.
2. Napaporn Sriden, Sandra Cortijo, Matt Box, Philip A Wigge, Varodom Charoensawan. Regulation of Temperature Transcriptomes in Hybrid Plants. Siriraj International Conference in Medicine and Public Health. June 2016.

W514–W521 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* *Published online 29 April 2016*
doi: 10.1093/nar/gkw314

MANORAA (Mapping Analogous Nuclei Onto Residue And Affinity) for identifying protein–ligand fragment interaction, pathways and SNPs

Duangrudee Tanramluk[1,2,*], Lalita Narupiyakul[2,3], Ruj Akavipat[2,4], Sungsam Gong[5] and Varodom Charoensawan[2,6]

[1]Institute of Molecular Biosciences, Mahidol University, Salaya, Nakhon Pathom 73170, Thailand, [2]Integrative Computational BioScience (ICBS) Center, Mahidol University, Salaya, Nakhon Pathom 73170, Thailand, [3]Department of Computer Engineering, Faculty of Engineering, Mahidol University, Salaya, Nakhon Pathom 73170, Thailand, [4]Department of Computer Science, Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand, [5]Department of Obstetrics and Gynaecology, University of Cambridge, The Rosie Hospital, Cambridge CB2 0SW, UK and [6]Department of Biochemistry, Faculty of Science, Mahidol University, Ratchathewi, Bangkok 10400, Thailand

Received February 8, 2016; Revised April 7, 2016; Accepted April 13, 2016

ABSTRACT

Protein–ligand interaction analysis is an important step of drug design and protein engineering in order to predict the binding affinity and selectivity between ligands to the target proteins. To date, there are more than 100 000 structures available in the Protein Data Bank (PDB), of which ∼30% are protein–ligand (MW below 1000 Da) complexes. We have developed the integrative web server MANORAA (Mapping Analogous Nuclei Onto Residue And Affinity) with the aim of providing a user-friendly web interface to assist structural study and design of protein–ligand interactions. In brief, the server allows the users to input the chemical fragments and present all the unique molecular interactions to the target proteins with available three-dimensional structures in the PDB. The users can also link the ligands of interest to assess possible off-target proteins, human variants and pathway information using our all-in-one integrated tools. Taken together, we envisage that the server will facilitate and improve the study of protein–ligand interactions by allowing observation and comparison of ligand interactions with multiple proteins at the same time. (http://manoraa.org).

INTRODUCTION

Understanding protein–ligand interaction is crucial for drug discovery research, as it defines the binding affinity, steric complementarity of the surface and pharmacophoric patterns of the compound to the target protein. Favorable ligand interactions with protein such as suitable polar groups counterparts and proper hydrogen bonding partners are crucial for the ligand design process and the imperfect fit between the protein and the ligand will result in decreased binding affinity (1). A number of tools is available for visualizing and analyzing protein–ligand interaction; however, only few can provide comprehensive information such as verified binding affinity, and couple the results with the ligand interaction visualization available for multiple protein comparison in the same place (2). By understanding the favorable interactions between the target protein and a ligand of interest, one can start to rationalize drug design strategy and make the protein engineering possible by strengthening preferred interactions for instance.

To date, there are more than 100 000 structures in the Protein Data Bank (PDB) (3). However, it is not always straightforward to harness all the relevant information from the PDB. Querying the substructure of the ligands to return multiple molecular interactions that are available in the PDB can take a considerable amount of time as one normally goes through a series of non-intuitive steps. After multiple protein–ligand structures are retrieved, the comparison can be complicated and time-consuming, especially when the structures contain a large amount of protein–ligand interactions from multiple contacts points, which normally have to be investigated individually and manually. To the best of our knowledge, there is no existing tool specifically designed for comparative analysis of protein–ligand interactions in multiple structures at the time.

Two of the most popular tools for searching molecular interactions in the binding sites are Relibase (4) and PDBeMotif (5). Both tools are restricted to the structures in the PDB and are often used to show the distribution of protein–ligand binding patterns in the PDB as a whole. Other tools such as PLIP (6) are also available for investigating protein

*To whom correspondence should be addressed. Tel: +66 2 441 9003 (Ext. 1211); Fax: +66 2 9428499; Email: duangrudee.tan@mahidol.ac.th

ligand interactions and visualization; however, the users cannot obtain knowledge of preferred interactions easily because it is dedicated to visualization and does not allow sorting by binding affinities, or viewing multiple proteins structures that bind to the same ligand in the same panel. The PLIC database (7) provides protein–ligand interaction clusters and also other related binding site information, and also has a superposition panel based on the clustering of similar binding sites. However, the ligand superposition is performed as a whole molecule, not based on the equivalent substructures, and hence, it is difficult to directly relate that information to the change in the binding affinities. WONKA (8) on the other hand, can offer observation from multiple structures but it requires the users to supply the set of superposed proteins with their equivalent amino acids renamed to the corresponding residue numbers. PoSSuM (9) aims to detect similar small molecule binding pockets; however, the overall similarity between pockets do not guarantee the same ligand binding pattern. A tool such as PLI (10) can also be used to find a particular ligand binding to a list of homologous proteins.

A direct query of ligand to the RCSB Protein Data Bank (3,11) returns the data retrieval in the form of PDB files and Jmol applet but does not provide ligand substructure analyses for multiple structures. The databases BindingDB (12) and Binding MOAD (13) emphasize the binding affinities data for further use such as for QSAR analysis (14), which does not offer the structural analysis of the binding site or the trend of binding affinity. In addition, these databases do not provide links from fragments to pathways or known human variants such as SNPs, a feature that will be useful for the drug design in the personalized medicine era.

To this end, we have developed the integrative web server MANORAA (Mapping Analogous Nuclei Onto Residue And Affinity) to facilitate understanding of ligand selectivity and promiscuity through the analysis of multiple protein structures on the web interface. It enables researchers to retrieve multiple chemical compounds and their binding partner proteins from the PDB, and compare and visualize the ligand-residue contact interactions all at the same time. Other useful functionalities include sorting of binding affinities of multiple proteins, as well as obtaining additional information such as protein functions, the species that a particular ligand is found in a complex, and the pathways that the ligand is found to take part in by linking to a pathway map such as KEGG (15), all in one place.

MANORAA: rationales, input and output

We built MANORAA with an aim to provide a user-friendly, one-stop service for ligand–protein interaction investigation. MANORAA was developed on top of CREDO, a database devoted specially to the protein–ligand interaction, which provides all pairwise atomic interaction contacts between ligand and proteins from the PDB in the form of a relational database (16,17). By filtering and ranking the interaction types in a systematic manner, the ligand contacts that are most important can be shown and can be related to the change in the binding affinities. This server provides integrated information about the target and off-target proteins interacting with the query ligand from the latest publicly available mirror of CREDO. MANORAA also provides protein–ligand binding affinity values from the Binding MOAD database (13) where the high quality binding affinity data are collected from literature. Importantly, the users can observe and compare the interaction by processing the ligand contact with multiple protein structures based on the complexes deposited to the PDB all at the same time. All queries to MANORAA start from one simple input page, and the results are provided in two output steps, as described here. We have extensively tested MANORAA on several common operating systems and web browsers (see Supplementary Data for details), and the most compatible browsers that we recommend are Windows version 7 or higher, and OSX Maverick or higher, and on Chrome 49 or higher, and Safari 9 or higher.

Input: chemical structure

The users can start with a ligand or part of a ligand of interest by providing one of the following as an input: (i) chemical name, (ii) SMILES expression, (iii) PDB ligand's 3-letter code or (iv) chemical structure (Figure 1). To facilitate generation of a SMILES expression, the MANORAA provides the SMILES lookup and then exports the SMILES to the chemical sketch panel as shown in Figure 1. The users can also select to create or edit a SMILES expression by drawing a chemical structure, or modify some parts and then import that sketch to a SMILES string before submission. From SMILES, the users can link to extended ligand names and compound bioactivity information via ChEMBL (18). The web server employs a JavaScript library called MarvinJS from ChemAxon to achieve the task. MarvinJS provides an HTML5-based user interface for chemical drawing, which allows the users to have an interactive interface without the need to install any additional plug-in.

Output 1: list of proteins interacting with a queried ligand

Once the users submit a ligand, a list of PDB entries that contain the submitted ligand will be returned. Ligand(s) of similar chemical structure and their target proteins will be returned if any part of the molecule matches with the SMILES input fragment. For example, Figure 2 shows a table of PDB entries interacting with a ligand 'STU' (Staurosporine). The binding affinity values, taken from Binding MOAD (13), are provided to help prioritize target proteins as they imply the binding strengths between the query ligand and the targets. For each entry, the following external information is also provided: (i) the pathway information from KEGG (15), (ii) the protein information from UniProt (19), (iii) the amino acid variants from SAMUL (20) and (iv) variants, isoforms and genomic context, protein/RNA baseline expression, gene ontology from the Centre for Therapeutic Target Validation (Open Targets, https://www.targetvalidation.org/). The server links crystallographic structures of protein–ligand interaction to related biochemical pathways via UniProt ID to KEGG ID mapping. The user can also link the proteins of interest to known human variants such as SNPs in the coding regions via the SAMUL web server (20). This essentially allows researchers to predict whether the candidate ligands will have a tendency to bind proteins with different annotated SNPs in the

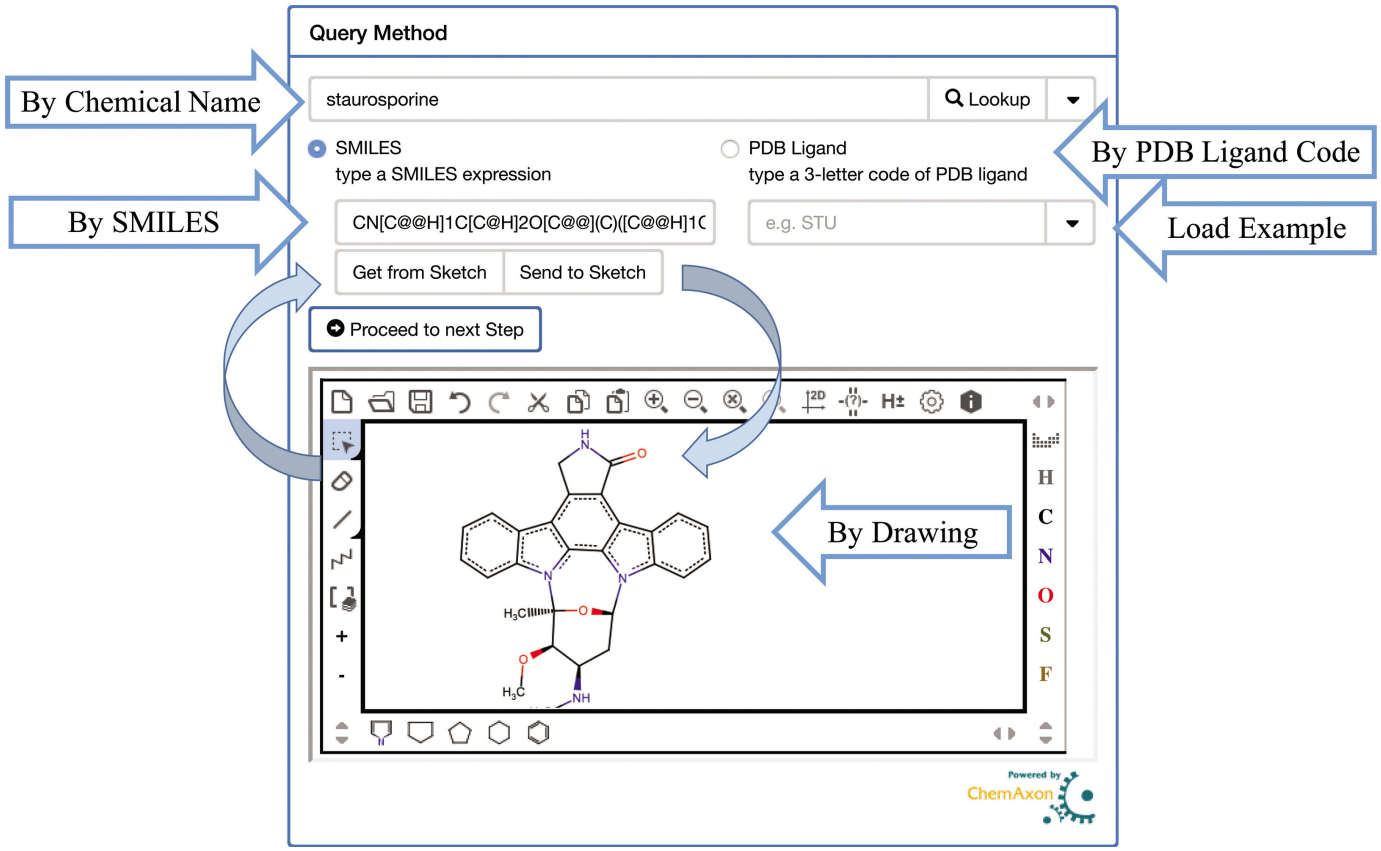


Figure 1. Input panel for ligand fragment via chemical names, SMILES expression, PDB ligand's three-letter code or chemical structure drawing.

coding regions, an important step of drug design in the personalized medicine era. These results can be exported as a CSV table.

This result page serves as an input form for the next step, which is to visualize three-dimensional (3D) structures of protein–ligand interactions based on the selection of (i) atom of interest in the ligand, and (ii) the target PDB chains. To aid this process, the MANORAA web server shows the ligand chemical structure so that users can pick up atoms of their interest interactively. By default, all the heteroatoms of ligands and PDB chains where the binding affinity are available are pre-selected.

Output 2: visualization of ligand–protein interactions

Once the users select ligand atoms and associated protein chains of interest, MANORAA will connect to CREDO to obtain interacting partner proteins of each ligand and PDB pair, grouped by nine interaction types and highlight them by different colors. The available interaction types are aromatic, hydrogen bond, ionic interaction, covalent bond, metal complex, carbonyl interaction, halogen bond, hydrophobic interaction and van der Waals clash (as shown in Figure 3 with criteria in Supplementary Data). The server will then rank the most important contact based on the shortest distances of unique interaction found for every atom per amino acid residue. JSmol (21), a JavaScript framework based on HTML5 for displaying interactive 3D molecular structures, has been employed into the user interface to enable the users to toggle display of the interaction partners in the 3D viewing panel. Display of ligand–protein interaction at each residue can be obtained by clicking the loading button of each PDB IDs, then choosing the residue name of interest. This step allows the user to have full control on what part of the chemical structure that they want to focus on. The results can be revisited using a unique URL provided. The list of target proteins and contact residues can be printed as PDF file together with the protein structures which can be saved from JSmol. Note that additional technical details of the web server can be found as Supplementary Data.

To assist the first-time users, we have provided comprehensive step-by-step tutorial, demo video and sample pages on the web server. Here, we also provide two examples of how MANORAA can be employed to assist real-world drug design and protein engineering research.

Making use of MANORAA in ligand–protein interaction studies

Case study 1: the trend of interaction observed in N4 of STU interacting with the kinase family. To illustrate the use of MANORAA and its features, here we use our previous comprehensive study on staurosporine's binding strength as an example (22). The study demonstrated that staurosporine's strength of interaction with kinase depends on the number and the orientation of hydrogen bonds and ionic interactions made around the N4 atom of staurosporine.

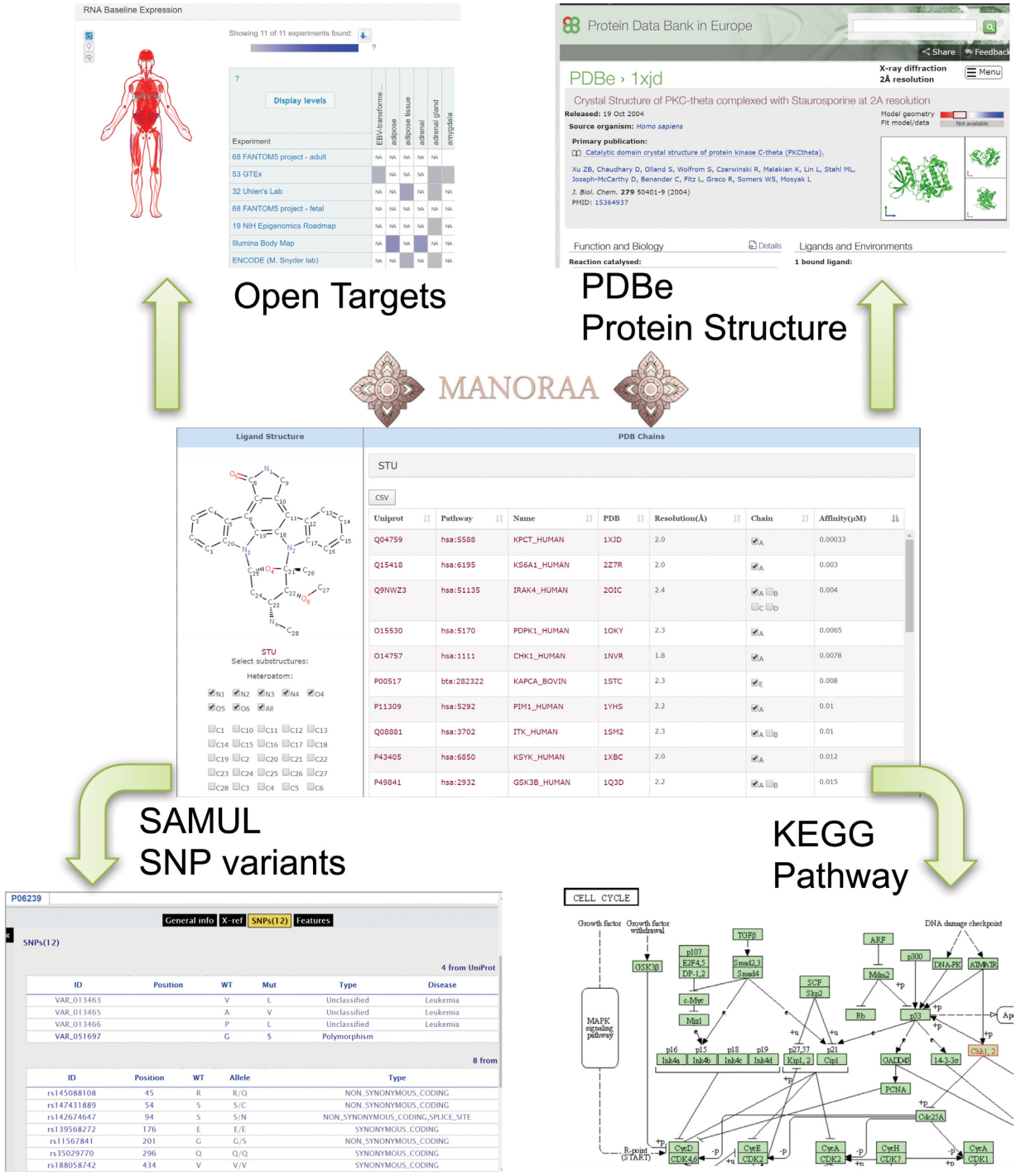
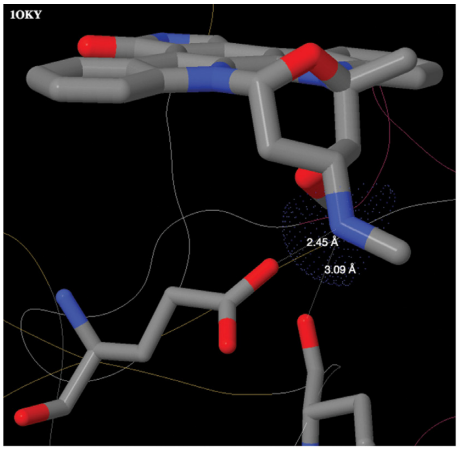


Figure 2. MANORAA provides integrated analyses for ligand–protein interactions, linking structural biology to genomics, pathways and target information. Middle inset: the ligand panel on the left shows a chemical fragment with information of their protein binding partners on the right. This allows users to make a query on their molecular interactions with available binding affinity, and obtain additional information about the target proteins/genes. Top left: target gene information such as baseline expression is provided via the Open Targets project, using UniProt name. Top right: link to the protein structure information via PDBe. Bottom left: SNP information via SAMUL. Bottom right: KEGG pathway where the protein/gene of interest is highlighted. The results can be sorted by the proteins' name, resolution, binding affinity and can be saved to a CSV file on the top of the page as needed.



Interaction Types		
Aromatic	Hydrogen Bond	Ionic
Covalent	Metal Complex	Carbonyl
Halogen Bond	Hydrophobic	vdW Clash

Interaction of Ligand: STU
Substructures: N4

LIGAND & PDB	Interacting partners		
1XJD & STU [STU] 200A & 1XJD:A **Protein kinase C theta type** from KPCT_HUMAN, Affinity=0.00033μM	HBOND	[ASP]508:A	Tight
	VDW	[ASP]508:A	
2Z7R & STU [STU] 1A & 2Z7R:A **Ribosomal protein S6 kinase alpha-1** from KS6A1_HUMAN, Affinity=0.003μM	HBOND	[ASP]148:A	
	IONIC	[ASP]148:A	
1OKY & STU [STU] 1368A & 1OKY:A **3-phosphoinositide-dependent protein kinase 1** from PDPK1_HUMAN, Affinity=0.0065μM	HBOND	[GLU]166:A [GLU]209:A	
	IONIC	[GLU]166:A	
	VDW	[GLU]166:A	
1NVR & STU [STU] 400A & 1NVR:A **Serine/threonine-protein kinase Chk1** from CHK1_HUMAN, Affinity=0.0078μM	HBOND	[GLU]134:A [GLU]91:A	
	IONIC	[GLU]91:A	
	VDW	[GLU]134:A [GLU]91:A	
1STC & STU [STU] 351E & 1STC:E **cAMP-dependent protein kinase catalytic subunit alpha** from KAPCA_BOVIN, Affinity=0.008μM	HBOND	[GLU]127:E [GLU]170:E	Medium
	IONIC	[GLU]127:E	
	VDW	[GLU]127:E [GLU]170:E	
1YHS & STU [STU] 306A & 1YHS:A **Proto-oncogene serine/threonine-protein kinase pim-1** from PIM1_HUMAN, Affinity=0.01μM	HBOND	[ASP]128:A [GLU]171:A	
	IONIC	[ASP]128:A	
	VDW	[ASP]128:A [GLU]171:A	
1SM2 & STU [STU] 301A & 1SM2:A **Tyrosine-protein kinase ITK/TSK** from ITK_HUMAN, Affinity=0.01μM	HBOND	[ARG]486:A	
	VDW	[ARG]486:A	
1XBC & STU [STU] 1A & 1XBC:A **Tyrosine-protein kinase SYK** from KSYK_HUMAN, Affinity=0.012μM	HBOND	[ARG]498:A	
1Q3D & STU [STU] 451A & 1Q3D:A **Glycogen synthase kinase-3 beta** from GSK3B_HUMAN, Affinity=0.015μM			Loose
1U59 & STU [STU] 100A & 1U59:A **Tyrosine-protein kinase ZAP-70** from ZAP70_HUMAN, Affinity=0.0558μM	HBOND	[ARG]465:A	
	VDW	[ARG]465:A	
3HMO & STU [STU] 1A & 3HMO:A **Dual specificity protein kinase TTK** from TTK_HUMAN, Affinity=0.102μM	IONIC	[ASP]608:A	

Figure 3. Display of the ligand three-letter code STU (staurosporine) interacting with several proteins in the kinase family. The top row shows interaction with the lowest binding affinity value, which means staurosporine can tightly bind (0.00033 μM) via aspartate, and there are at most two hydrogen bonds (cyan) or ionic interaction (magenta) in combination. The second group has binding affinity values between 0.0065 to 0.010 μM and has three interactions that are either ionic or hydrogen bonds. All the others with binding affinity value more than 0.010 μM have one hydrogen bond, or ionic interaction, or no interaction at all.

For the worst STU binding to kinase cases, N4 from PDB IDs: 1XBC, 1U59, 3HMO have only one bond in the non-preferable orientation while the one from 1Q3D does not have any hydrogen bonding or ionic interaction at all. These two structural observations imply poor binding affinity and hence they are present in the group of large binding affinity values. In contrast, 1OKY, 1NVR, 1STC and 1YHS have better binding affinities due to better hydrogen bonding and ionic interaction in both the structural orientation sense and also the amount of bonds they made which are at least two hydrogen bonds plus one ionic interaction. Note that binding via aspartate makes the interaction tighter than binding via arginine as seen in 1XJD and 2Z7R and hence shown with the lowest Ki or tightest binding affinities from the Binding MOAD database. These types of analysis will be beneficial to drug design because we know which part of the ligand is the major determinant of the binding affinities and what amino acids facilitate those preferred interactions. With MANORAA, these processes can be performed all in one place (Figure 3).

Case study 2: how MANORAA can be powerful for analyzing ligand and its binding protein with known SNPs that can be linked to diseases. Trifluoperazine (PDB ligand code TFP) was originally identified as an antipsychotic drug used in the treatment of schizophrenia, via blocking D2 dopaminergic receptors in the brain. However, trifluoperazine has recently been repurposed to inhibit the growth of cancer stem cells via its function as a calmodulin inhibitor, but the inhibitory mechanism is unclear (23,24). Using MANORAA, we can demonstrate that this ligand can also bind with many target proteins such as the placental calcium binding protein (S100-A4), calmodulin and troponin C (in a way, 'off-targets' to dopaminergic receptors). The result shows that the N3 atom of TFP interact with the placental calcium binding protein (S100-A4) via one hydrogen bond, or one optional ionic interaction in the cases of bovine calmodulin and human troponin C, suggesting the importance of this atom for all the proteins that bind to this small molecule. On the other hand, the N2 atom of TFP forms two ionic interactions with both bovine calmodulin structures but does not make any significant interaction for either S100-A4 protein or human troponin C. This kind of information can be useful for designing the selectivity of the drug.

Furthermore, MANORAA (via SAMUL) also reveals two SNPs in the human calmodulin gene that have been associated with ventricular tachycardia, a common side effect from trifluoperazine use. MANORAA also provides a list of multiple bovine calmodulin with crystallographic structures that harbor the ligand, enabling researchers to explore the effect of amino acid changes to affect ligand–protein interaction. This demonstrates how MANORAA can be used for an initial assessment of drug repurposing results. It should be noted that our data relies on crystallographic structures deposited to the PDB at the time. Saying that, MANORAA provides another way to make use of the growing PDB by linking the structures to human genome variations.

DISCUSSION AND FUTURE DIRECTIONS

Thanks to technological advances in crystallography and other methods for determining structures of biological molecules, the bottleneck of structural biology is now shifting from obtaining the structures to interpreting and linking them to other biological information such as pathways and genomic variants. With the wealth of information on ligand–protein interactions from publicly available databases such as PDB, it is now possible to perform a comparative study of multiple ligands and proteins (or drug candidate compounds and target proteins) at the same time. MANORAA has been established to facilitate these processes all in one place.

The web server has a number of useful features that assist the investigation of ligand–protein binding specificity, biological pathways the proteins are involved in and known human variants in the coding regions of the proteins. As depicted by Böhm, ligand with poor binding affinity is caused by missing crucial active site interactions in comparison with other tight binding ligands (1). Our service allows the user to compare the ligand's binding affinities with the numbers and the types of interactions that the ligand makes with multiple proteins, which should be useful for users to identify the key residues of proteins and the atoms of ligands in order to manipulate the interaction strength.

Existing protein–ligand contact and interaction databases are required in order to expedite the process of calculating and classifying the molecular interaction on the fly. To this end, we make use of the CREDO backend, and provide links to SAMUL (20), UniProt (19), PDBe (5), PDBsum (25) and KEGG (15). Note that the total size of calculated interaction for CREDO databases alone, including all the structures in the PDB, is very large (72 GB), but that would allow the interactions to be observed almost instantaneously. The graphical representation on JSmol (21) allows the users to view multiple structures with the fragment in the same window. The color highlight of the protein–ligand interaction that links to the JSmol structure visualization panel in real-time allows robust ligand interaction identification so that the researchers can relate the knowledge of the binding affinity value to the missing or occurring interaction by themselves. MANORAA employs a support responsive design, which means the output structures can be visualized without distortion even on a tablet or mobile phone. Another unique feature of MANORAA is its multiple-structures visualization panel with multiple loading buttons. The users can observe multiple structures one at a time and progress to each one to get an impression of the whole set of proteins that interact with this particular ligand fragment, and could identify the amino acid residue or atoms of a ligand that can be modified to fine-tune the ligand–protein binding interaction.

The main strengths of MANORAA over other previously available web servers aforementioned includes its flexibility of analyzing multiple experimentally verified ligand–protein interactions at the time, using its user-friendly and fast responsive interface. PLIP (6), for instance, focuses on the visualization of one structure at a time, while PoSSuM (9) provides the superposition and makes a comparison be-

tween residues surrounding a protein pair, rather than giving details of the type of residues in contact. Even though there are a number of tools that allow multiple structure observation, the interactions are not dissected to different chemical interaction types and do not provide visualization panel for the users to drill down to the level of ligand substructure interactions.

Looking ahead, we aim to routinely maintain the server and add new functionalities, which will be managed by a programmer dedicated to MANORAA's development and a multidisciplinary team. For instance, we have been developing a new algorithm to show gradient color of atomic position conservation. This will allow us to show position-specific interaction by highlighting the active site based on the percent conservation of the atomic position surrounding the ligand substructure. For the time being, we have implemented this for 18 staurosporine superposed complexes as an example from our sample page (see Supplementary Data). In addition, protein ligand contacts will be updated for every newer release of CREDO. In addition, PDBe (5), PDBsum (25), CACTVS (26,27), ChEMBL (18), KEGG (15), Open Targets (https://www.targetvalidation.org/), UniProt (19) and SAMUL (20) are accessed in real-time through their websites, hence the results shown will always be the most updated.

With MANORAA, the chemical fragments that have an influence on different pathways in different organisms can open the door for a more robust and insightful analysis for the study of multi-target drug design, species selectivity, off-target inhibition causing drug side effect problems. We envisage that MANORAA will provide a missing link between structural biology, systems biology and genetics information by one central concept surrounding the ligand's chemical structure to assist drug discovery and the probe molecules community.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Sir Tom Blundell, Dr Adrian Schreyer and Dr Bernardo Ochoa who provided the CREDO database and M.R. Jisnuson Svasti who supported the initial phase of the MANORAA server. We thank Pisitsak Sriwimol and Sakao Phoochaijaroen for their technical supports. We also thank Dr Bhoom Suktitipat and Lect. Stephen Smith for useful comments on the manuscript. Marvin JS 15.2.23, 2015, ChemAxon (http://www.chemaxon.com), was used for drawing and displaying chemical structures and substructures.

FUNDING

Mahidol University; Thailand Research Fund via Institute of Molecular Biosciences and Faculty of Science, Mahidol University [MRG5480208 to D.T., TRG5880067 to V.C]; Crown Property Bureau Foundation via the Integrative Computational BioScience (ICBS) center, Mahidol University. Funding for open access charge: Thailand Research Fund [IRG5780009]; Mahidol University Research Grant.

Conflict of interest statement. None declared.

REFERENCES

1. Böhm,H.-J. (2003) *Protein–Ligand Interactions—From Molecular Recognition to Drug Design*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
2. Inhester,T. and Rarey,M. (2014) Protein–ligand interaction databases: advanced tools to mine activity data and interactions on a structural level. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **4**, 562–575.
3. Rose,P.W., Prlic,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
4. Hendlich,M., Bergner,A., Gunther,J. and Klebe,G. (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.*, **326**, 607–620.
5. Velankar,S., Alhroub,Y., Best,C., Caboche,S., Conroy,M.J., Dana,J.M., Fernandez Montecelo,M.A., van Ginkel,G., Golovin,A., Gore,S.P. *et al.* (2011) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, **40**, D445–D452.
6. Salentin,S., Schreiber,S., Haupt,V.J., Adasme,M.F. and Schroeder,M. (2015) PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.*, **43**, W443–W447.
7. Anand,P., Nagarajan,D., Mukherjee,S. and Chandra,N. (2014) PLIC: protein–ligand interaction clusters. *Database*, bau029.
8. Bradley,A.R., Wall,I.D., von Delft,F., Green,D.V., Deane,C.M. and Marsden,B.D. (2015) WONKA: objective novel complex analysis for ensembles of protein-ligand structures. *J. Comput. Aided Mol. Des.*, **29**, 963–973.
9. Ito,J.-i., Ikeda,K., Yamada,K., Mizuguchi,K. and Tomii,K. (2014) PoSSuM v.2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs. *Nucleic Acids Res.*, **43**, D392–D398.
10. Gallina,A.M., Bisignano,P., Bergamino,M. and Bordo,D. (2013) PLI: a web-based tool for the comparison of protein-ligand interactions observed on PDB structures. *Bioinformatics*, **29**, 395–397.
11. Rose,P.W., Bi,C., Bluhm,W.F., Christie,C.H., Dimitropoulos,D., Dutta,S., Green,R.K., Goodsell,D.S., Prlic,A., Quesada,M. *et al.* (2013) The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.*, **41**, D475–D482.
12. Liu,T., Lin,Y., Wen,X., Jorissen,R.N. and Gilson,M.K. (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.*, **35**, D198–D201.
13. Benson,M.L., Smith,R.D., Khazanov,N.A., Dimcheff,B., Beaver,J., Dresslar,P., Nerothin,J. and Carlson,H.A. (2008) Binding MOAD, a high-quality protein–ligand database. *Nucleic Acids Res.*, **36**, D674–D678.
14. Hansch,C. and Silipo,C. (1974) Quantitative structure-activity relation of reversible dihydrofolate reductase inhibitors. Diaminotriazines. *J. Med. Chem.*, **17**, 661–667.
15. Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
16. Schreyer,A. and Blundell,T. (2009) CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol. Drug Des.*, **73**, 157–167.
17. Schreyer,A.M. and Blundell,T.L. (2013) CREDO: a structural interactomics database for drug discovery. *Database*, bat049.
18. Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Kruger,F.A., Light,Y., Mak,L., McGlinchey,S. *et al.* (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.
19. Bateman,A.M.M., O'Donovan,C., Magrane,M., Apweiler,R., Alpi,E., Antunes,R., Arganiska,J., Bely,B., Bingley,M., Bonilla,C. *et al.* (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
20. Gong,S., Worth,C., Cheng,T.K. and Blundell,T. (2011) Meet me halfway: when genomics meets structural bioinformatics. *J. Cardiovas. Transl. Res.*, **4**, 281–303.
21. Hanson,R.M. (2010) Jmol – a paradigm shift in crystallographic visualization. *J. Appl. Cryst.*, **43**, 1250–1260.
22. Tanramluk,D., Schreyer,A., Pitt,W.R. and Blundell,T.L. (2009) On the origins of enzyme inhibitor selectivity and promiscuity: a case

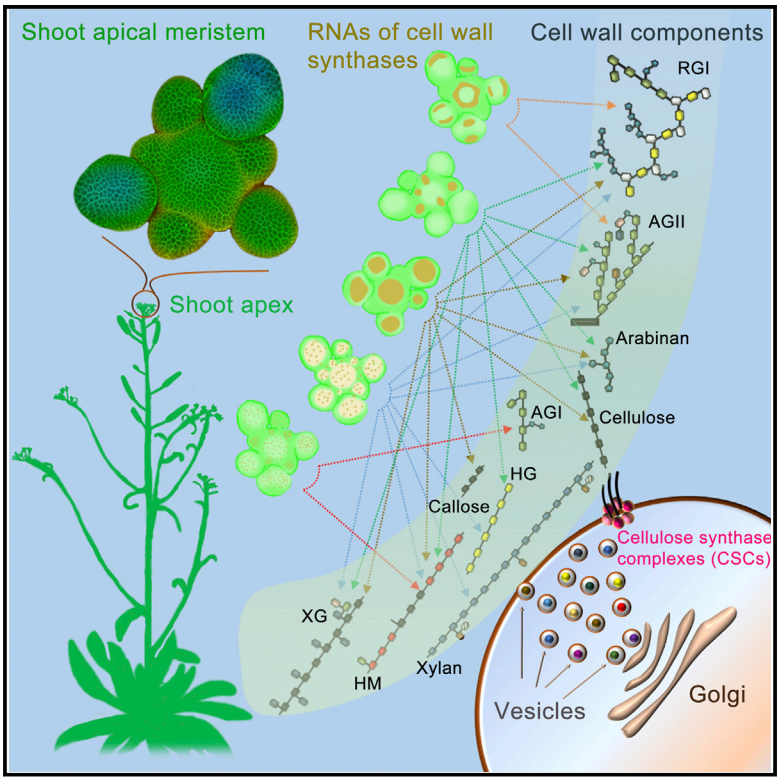
study of protein kinase binding to staurosporine. *Chem. Biol. Drug Des.*, **74**, 16–24.
23. Pulkoski-Gross,A., Li,J., Zheng,C., Li,Y., Ouyang,N., Rigas,B., Zucker,S. and Cao,J. (2015) Repurposing the antipsychotic trifluoperazine as an antimetastasis agent. *Mol. Pharmacol.*, **87**, 501–512.
24. Yeh,C.-T., Wu,A.T.H., Chang,P.M.H., Chen,K.-Y., Yang,C.-N., Yang,S.-C., Ho,C.-C., Chen,C.-C., Kuo,Y.-L., Lee,P.-Y. *et al.* (2012) Trifluoperazine, an antipsychotic agent, inhibits cancer stem cell growth and overcomes drug resistance of lung cancer. *Am. J. Respir. Crit. Care Med.*, **186**, 1180–1188.
25. Laskowski,R.A., Hutchinson,E.G., Michie,A.D., Wallace,A.C., Jones,M.L. and Thornton,J.M. (1997) PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.*, **22**, 488–490.
26. Ihlenfeldt,W., Takahashi,Y., Abe,H. and Sasaki,S. (1994) Computation and management of chemical properties in CACTVS: An extensible networked approach toward modularity and compatibility. *J. Chem. Inf. Comput. Sci.*, **34**, 109–116.
27. Ihlenfeldt,W.D., Voigt,J.H., Bienfait,B., Oellien,F. and Nicklaus,M.C. (2002) Enhanced CACTVS browser of the Open NCI Database. *J. Chem. Inf. Comput. Sci.*, **42**, 46–57.

Article

Current Biology

Regulation of Meristem Morphogenesis by Cell Wall Synthases in *Arabidopsis*

Graphical Abstract



Highlights

- Primary cell walls of a plant shoot meristem and flowers were analyzed
- Polysaccharide content was shown to vary depending on cell type and cell-cycle stage
- Genes encoding wall biosynthetic enzymes were matched to polysaccharide pattern
- The *CSLD* gene family plays a key role in meristem morphology, growth, and patterning

Authors

Weibing Yang, Christoph Schuster, Cherie T. Beahan, ..., Monika S. Doblin, Raymond Wightman, Elliot M. Meyerowitz

Correspondence

msdoblin@unimelb.edu.au (M.S.D.), raymond.wightman@slcu.cam.ac.uk (R.W.), meyerow@caltech.edu (E.M.M.)

In Brief

Yang et al. characterize the changes in plant cell walls and the expression of the genes for the enzymes that create them in the *Arabidopsis* shoot apex. They find that different enzymes build new walls and maturing walls, and that wall components and the gene expression that underlies their synthesis are highly regulated throughout development.

Yang et al., 2016, Current Biology *26*, 1–12
June 6, 2016
http://dx.doi.org/10.1016/j.cub.2016.04.026

CellPress

Please cite this article in press as: Yang et al., Regulation of Meristem Morphogenesis by Cell Wall Synthases in *Arabidopsis*, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.04.026

Current Biology
Article

Regulation of Meristem Morphogenesis by Cell Wall Synthases in *Arabidopsis*

Weibing Yang,[1] Christoph Schuster,[1] Cherie T. Beahan,[2] Varodom Charoensawan,[1,3,4] Alexis Peaucelle,[1,5] Antony Bacic,[2] Monika S. Doblin,[2,*] Raymond Wightman,[1,*] and Elliot M. Meyerowitz[1,6,*]
[1]Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge CB2 1LR, UK
[2]ARC Centre of Excellence in Plant Cell Walls, School of BioSciences, The University of Melbourne, Parkville, VIC 3010, Australia
[3]Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
[4]Integrative Computational BioScience Center, Mahidol University, Nakhon Pathom 73170, Thailand
[5]Institut Jean-Pierre Bourgin, UMR 1318, Institut National pour la Recherche Agronomique-AgroParisTech, Saclay Plant Science, Route de St-Cyr, Versailles 78026, France
[6]Howard Hughes Medical Institute and Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
*Correspondence: msdoblin@unimelb.edu.au (M.S.D.), raymond.wightman@slcu.cam.ac.uk (R.W.), meyerow@caltech.edu (E.M.M.)
http://dx.doi.org/10.1016/j.cub.2016.04.026

SUMMARY

The cell walls of the shoot apical meristem (SAM), containing the stem cell niche that gives rise to the above-ground tissues, are crucially involved in regulating differentiation. It is currently unknown how these walls are built and refined or their role, if any, in influencing meristem developmental dynamics. We have combined polysaccharide linkage analysis, immuno-labeling, and transcriptome profiling of the SAM to provide a spatiotemporal plan of the walls of this dynamic structure. We find that meristematic cells express only a core subset of 152 genes encoding cell wall glycosyltransferases (GTs). Systemic localization of all these GT mRNAs by in situ hybridization reveals members with either enrichment in or specificity to apical subdomains such as emerging flower primordia, and a large class with high expression in dividing cells. The highly localized and coordinated expression of GTs in the SAM suggests distinct wall properties of meristematic cells and specific differences between newly forming walls and their mature descendants. Functional analysis demonstrates that a subset of *CSLD* genes is essential for proper meristem maintenance, confirming the key role of walls in developmental pathways.

INTRODUCTION

Plant biomass, our only renewable bioresource, is largely composed of cell walls. The primary plant cell wall is a complex composite made up almost entirely of polysaccharides (90%–95%), with some (glyco)proteins (5%–10%) [1–4]. It serves to provide strength and mechanical support to plant tissues and provides resistance to the high turgor pressure inside each cell. Local reinforcement coupled with wall loosening, achieved by rapid remodeling, permits not only growth but also the generation of a variety of cell shapes, ranging from the cylindrical cells of the epidermis and endodermis of the root to more complex shapes such as those of the leaf epidermal pavement cells [5]. All growing cells contain a primary wall, and further specialization is observed in certain cell types during tissue differentiation [1, 6].

The shoot apical meristem (SAM) is a dome-shaped structure that contains the above-ground stem cell pool, slowly proliferating cells that are found at the top of the dome within a region termed the central zone (CZ). The progeny of these cells are gradually displaced to the peripheral zone (PZ), where cells grow and divide at a higher rate [7]. Auxin maxima within the PZ determine the sites of either flower or leaf primordial initiation, characterized by maximal rates of cell expansion [8, 9]. It has been proposed that the dome-shaped structure is maintained through a feedback loop whereby the stress patterns, dictated by tissue geometry, influence the organization of the cytoskeleton and reinforcements of the cell wall [10]. This loop affects auxin movements around the SAM via changes in the polarity of the PIN1 auxin transporter [11]. Differences in cell wall rigidity appear to demarcate the different functional domains [12], with the CZ being 3-fold stiffer (having a higher elastic modulus) than faster-growing cells at the flanks of the SAM [13, 14]. Thus, local differences exist in walls, and these result from, and contribute to, developmental dynamics.

The primary wall of most flowering plants consists typically of cellulose, non-cellulosic polysaccharides, pectins, and glycoproteins/proteoglycans. Whereas cellulose is an unsubstituted homopolymer of glucose (Glc), most polysaccharides have backbone substitutions ranging from a small number of sugar residues (e.g., xyloglucan; XG) to the very complex branching patterns observed in some pectic polysaccharides, such as arabinogalactan (AG), rhamnogalacturonan (RG), and proteoglycans (AGPs). Both glycan backbone chain elongation and substitutions are carried out by polysaccharide synthases/glycosyltransferase (GT) enzymes, classified into families and subfamilies based on phylogenetic similarities. *Arabidopsis thaliana* possesses on the order of 560 GTs, and many of these are expected to be involved in cell wall synthesis (CAZy; http://www.cazy.org).

Despite the central role played by the cell wall in the SAM and subsequent development, very little information exists as to its

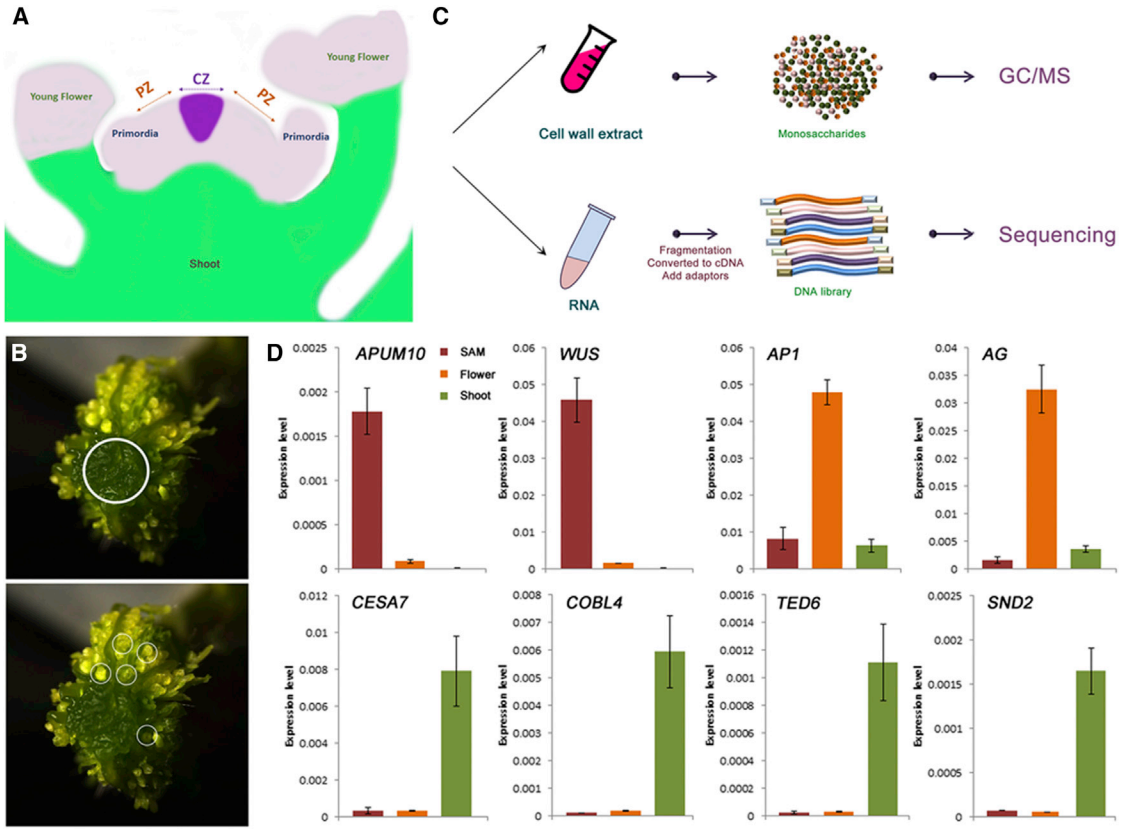


Figure 1. Dissection of Meristematic Cells for Wall Analysis

(A) Schematic representation of the organization of the *Arabidopsis* shoot apex. Meristematic cells (purple) in both the shoot apical meristem (SAM) and flower primordia were collected for analysis. CZ, central zone; PZ, peripheral zone.

(B and C) Schematic flowchart displaying the strategy used in this study. (B) The meristematic cells carefully dissected from *clv3-2* SAM (top panel, circled) and flower primordia (before Stage 6 according to [20] and shown as circles in the bottom panel) were used for RNA extraction and cell wall preparation followed by linkage analysis (C).

(D) The expression of genes specific to SAM (*APUM10* and *WUS*), flower (*AP1* and *AG*), and shoot vasculature (*CESA7*, *COBL4*, *TED6*, and *SND2*) in dissected SAM and young flower. The shoot sample was included for comparison. The *Arabidopsis UBIQUITIN10* gene was used as an internal control. Shown are mean values from three replicates; error bars represent SD values.

See also Figure S1 and Table S3.

composition and that of the early-developing organs [14–18]. Here we have combined data from polysaccharide linkage analysis, steady-state mRNA quantification and localization, and polysaccharide immuno-labeling to build up a comprehensive picture of the construction of the SAM primary walls and their cognate biosynthetic GTs. We find that meristematic cell walls are constructed by a reduced subset of GTs with spatially and temporally regulated expression patterns. The data suggest a distinction between new walls formed at cell division and pre-existing walls, as well as a key role by a GT of the *CSLD* gene family, required for proper growth and maintenance of cell number in the SAM.

RESULTS

Isolating Shoot Meristematic Cells for Polysaccharide and Expression Profiling

To reveal the structure and synthesis of stem cell walls, we tried to harvest pure meristematic cells from the *Arabidopsis* shoot apex. Due to the small meristem size, it would require around 10^5 dissected SAMs to generate the 4 mg of fresh weight required for reliable polysaccharide linkage analysis. We thus looked to the enlarged, stem cell-enriched SAMs of the *clavata3* (*clv3*) mutant [19] (Figure S1). Meristems were carefully dissected from 60 *clv3-2* shoot apices, and young flowers were also included in our analysis to represent a mixed sample of meristematic, highly proliferative, and developing tissues (Figures 1A–1C) [20]. The samples were subjected to cell wall linkage analysis and RNA sequencing (Figure 1B). To evaluate the quality of these dissected samples (i.e., no contamination of developed shoot tissues in the meristematic cells), we performed qRT-PCR to check genes with specific expression pattern in different tissues. Consistently, *APUM10* and *WUS*, two SAM-enriched genes, showed high expression in the SAM sample, and floral organ-specific genes, *AP1* and *AG*, expressed highly in the flower sample. By contrast, the transcripts of genes involved in secondary cell wall biosynthesis and vascular formation, including *CESA7*, *COBL4*, *TED6*, and *SND2*, could barely be detected in the meristematic tissues (Figure 1D).

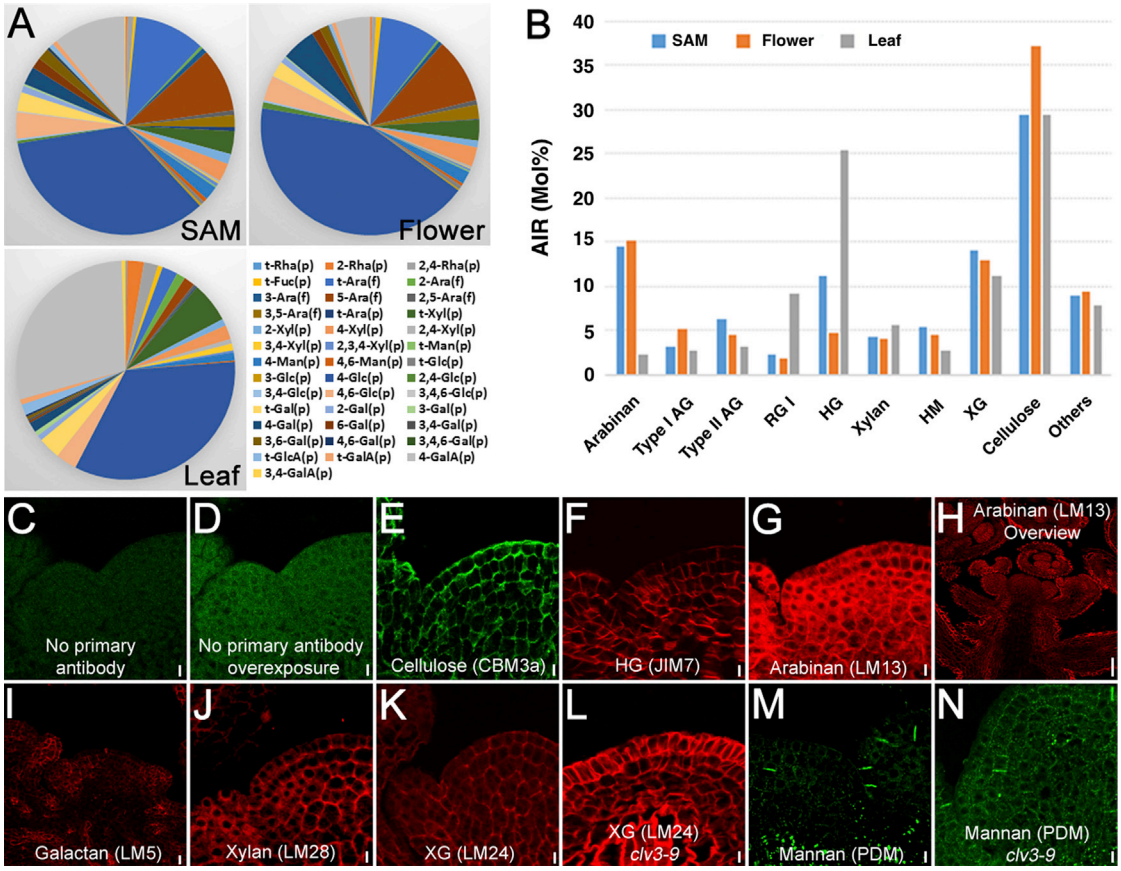


Figure 2. Composition and Spatial Distribution of Wall Components in the Shoot Apex

(A) Monosaccharide linkage composition of AIR cell wall preparations.

(B) Calculation of polysaccharide composition based on the monosaccharide linkage analysis shown in (A).

(C–N) Immunofluorescence labeling showing the spatial distribution of wall components. The longitudinal sections of the *Arabidopsis* shoot apex were incubated with cell wall antibody probes. No primary antibody (anti-rabbit secondary) control is shown in (C), and overexposed control is shown in (D). Scale bars, 5 μm (C–G and J–N), 20 μm (H), and 10 μm (I).

See also Table S1.

Composition and Spatial Organization of Cell Wall Components

The alcohol-insoluble residues (AIRs) from samples were analyzed to determine their polysaccharide composition by monosaccharide linkage analyses based upon a priori knowledge of the relative proportions of these linkages in a particular polysaccharide [21]. Most of the monosaccharides identified previously from leaf could also be detected in both meristem and flower, although the relative composition was different (Figure 2A). Polysaccharide calculations show the walls of meristematic cells to be composed of, in order of decreasing abundance, cellulose, pectin, and non-cellulosic polysaccharides (Figure 2B). A marked increase in the amount of homogalacturonan (HG), composed of linear chains of galacturonic acid (GalA) residues, was found in the SAM compared to flowers (11.3% versus 4.8%), whereas type I AG (3.3% versus 5.3%) and cellulose (29.4% versus 37.1%) exhibited decreases in the SAM. Compared to leaf, both the SAM and young floral tissue contained at least 3-fold less RGI, a 2- to 6-fold reduction in HG, and an approximately 6-fold increase in pectic arabinan. Total pectin content is reduced in the SAM and flowers compared to the level in leaves. Cellulose levels in the leaf (29.5%) were found to be nearly identical to the SAM (29.4%). Overall, the data show some marked differences in cell wall pectic and non-cellulosic composition between the SAM and its derivatives, the flower and leaf.

To reveal the spatial organization of cell wall structure in the SAM, we took advantage of immunohistochemistry using different antibodies or recombinant fluorescent proteins that recognize individual wall components. Consistent with cellulose being the major polysaccharide, CBM3a, which binds crystalline cellulose, was found to label all walls within the SAM. Labeling of walls across the apex by CBM3a was not uniform, suggesting either differences in the organization of wall polymers (thereby affecting accessibility of the probe to cellulose), variation in the degrees of cellulose crystallinity, or absolute abundance (Figure 2E). Similarly, inhomogeneous labeling was observed for JIM7 (HG), which might reflect some variation in the degree of HG methyl-esterification (Figure 2F). Detection of pectic α-1,5-arabinan by LM13 resulted in a complex pattern of labeled walls in both SAM and floral tissues (Figure 2G), and a very intense signal was seen in the epidermal cells of older developing tissues (Figure 2H). Labeling of β-1,4-galactan by LM5 was found predominantly in primordia

and the outer whorl of young developing flowers (Figure 2I), correlating closely with the double amount of 1,4-Gal in flowers compared to the SAM from linkage analysis (5.1% versus 2.7%; Table S1). Both heteroxylan (HX) and XG existed in all walls of meristematic cells, as revealed by LM28 and LM24 labeling (Figures 2J and 2K). However, the signals could be observed only after pectolyase treatment, suggesting that these polysaccharides are masked by pectin. Although most of the epitopes exhibited similar organization patterns in *clv3* compared to wild-type, XG seems to show distinct patterns, with a higher signal in L1 layer cells (Figure 2L). This implies a secondary effect of an enlarged CZ upon either XG synthesis or organization. In addition, we found that PDM, a mannan antibody that recognizes heteromannans (glucomannans and galactomannans), only labeled certain regions of cells that oriented anticlinally in upper cell layers but were situated either anticlinally or periclinally in lower cell layers (Figures 2M and 2N), reminiscent of the newly formed cross walls arising after cell-division events.

Analysis of the Transcriptome Reveals Differential Expression of a Subset of GT Genes in the SAM

With this insight into the cell wall composition of the SAM and young flowers, we could then examine the specific differential expression patterns of GTs in these tissues, because these enzymes determine, to a large extent, which wall polysaccharides and glycoproteins are being made. We performed RNA sequencing using the highly purified SAM tissues and young flowers that were equal to those used for linkage analysis. A third sample, consisting of pooled total RNA from several plant tissues, was also analyzed for comparison. Comparison of the RNA levels from our purified meristematic cells with previously published cell-type expression profiling data revealed high spatial sensitivity to detect genes expressed within particular tissues (see the Supplemental Experimental Procedures). For example, *APUM10/PUMLIO10*, a gene that is only expressed in the *CLV3* domain, was detected exclusively in the SAM samples (Figure S2A). Gene ontology (GO) analysis of differentially expressed genes revealed cell wall categories to be under-represented in the SAM and flower (Figure S2B). Further analysis found differences in GT transcript abundance between SAM, flower, and the whole-plant samples (Figure S2C), suggesting that meristematic cells employ only a subset of GT enzymes to build the wall. For example, in the GT8 subfamily of enzymes that give rise to GlcA substitution of xylan [22–24], only *GUX3* expression could be detected in the SAM and flower. Among the ten *FUT* genes responsible for XG and AGI synthesis, only three exhibited considerable expression in meristematic tissues. This contraction also extends to the protein backbones of proteoglycans, including those that make up AGPs and extensins, where only a subset of mRNAs are present in the SAM (Table S2). A literature search allowed the assignment of more gene families or subfamilies that encode cell wall-related GTs to polysaccharide linkages. A total of 152 SAM-expressed GTs (transcripts per million [TPM] value >1) were mapped to their respective polysaccharides/glycoproteins (Table S2).

Consistent with cellulose being the major polysaccharide, the core primary wall cellulose synthase subunits of the GT2-CESA family, encoded by *CESA1* and *CESA3*, are highly co-expressed (TPM values >100 for all samples), as observed in all primary wall tissues, and are found together with other *CESA* mRNAs (*CESA2*, *CESA5*, and *CESA6*) that encode proteins that provide the third component of the active complex [25], with both *CESA5* and *CESA6* exhibiting higher mRNA levels in flowers.

The pectin backbone (HG and RGI) (Figure 2B) is synthesized by GalATs and encoded by the GT8-GAUT subfamily [26], the transcripts of which are widely represented in the SAM. The proposed HG core complex is encoded by *GAUT7* and *GAUT1*, and these are both expressed (Table S2), consistent with their products forming a core biosynthetic complex [27]. Despite the low quantities of RGI in both the SAM and flower compared to leaf, expression of some potential RGI-associated GTs (*GATL3*, *5*, *6*, *8*, and *9*) were found at higher levels than for the pooled samples (Table S2). The absence of *XGD1* mRNA in the floral sample suggests that flowers either do not contain xylogalacturonan or that it is made by the product of the related *EMB175* gene.

XG biosynthesis has only one confirmed GT2 family member, namely *CSLC4*, responsible for backbone (1,4-Glc) synthesis [28]. Furthermore, all the genes encoding enzymes that make up the recently identified XG multi-protein complex, *CSLC4*, *XXT2*, *XXT5*, *MUR3* [29], and *XLT2* [30], are expressed in the SAM at levels comparable to the pooled sample (Table S2). *CSLC6* displays the highest expression of all *CSLC* subfamily members in the SAM; however, a function in XG assembly has not yet been demonstrated. The GT34 family contains seven members, of which data support five genes encoding XG xylosyltransferases (*XXT1–5*). There is no evidence that the remaining two, designated *GTL6* and *GTL7*, encode proteins with this type of GT activity [31]. *GTL6* is one of the most highly expressed GT genes in the SAM (mean TPM 210; Table S2). The gene product exhibits similarity to galacto(gluco)mannan galactosyltransferases (GMGTs) in other plants [32]. Indeed, recent data demonstrate that the gene, renamed *MUCI10*, encodes a GMGT involved in seed mucilage formation [33]. We have therefore assigned these two GMGTs to mannan as the enzymes that add the terminal Gal residue onto the 4-Man backbone, giving rise to the 1,4,6-Man residue that most likely represents substituted heteromannan. The backbone is synthesized by members of the CSLA family, where *CSLA2* transcripts are the most abundant (Table S2). *CSLA7*, essential for embryonic development [34], and *CSLA10* exhibit higher expression in the flower compared to other tissues (Table S2).

For HX, a proportion detected in the floral linkage analysis would be expected to come from the vascular-derived secondary walls; however, it is not expected that there would be the same in the SAM preparations, because mRNA transcripts from known genes involved in xylem and secondary wall formation have been shown either to be absent or at barely detectable levels (Figure 1D; Figure S2A). Although the key genes encoding the enzymes for 1,4-Xyl backbone synthesis have been shown to be *IRX9*, *IRX10*, and *IRX14* [35, 36], transcriptome data suggest apical enrichment of the corresponding paralog in each case (*IRX9L*, *IRX10L*, and *IRX14L*; Table S2). Of the secondary wall-associated members, only *IRX10* and *IRX14* are expressed in the SAM. Also, F8H rather than FRA8 is the likely GT involved in making the reducing-end sequence of the xylan chain.

Although the presence of a GT transcript is expected to correlate with its cognate polysaccharide, within the complex cell wall composite we might expect to see spatiotemporal correlation

between different polysaccharides reflected in the expression of associated GTs. Using collections of publicly available microarray expression data, a pairwise correlation coefficient network [37] was constructed from SAM GT transcripts higher than 10 TPM (Figure S2D). At the center of the network are the *CESA1*, *3*, and *6* genes that have a pairwise Pearson coefficient $r > 0.9$, consistent with the encoded proteins forming a primary wall cellulose synthase complex (CSC) found in diverse tissue types [25]. Figure S2D (boxed) shows a subnetwork that comprises genes with transcripts that peak in the young floral samples and where the core is occupied by *CSLC8*. Together with *CSLC5*, this may represent construction of an XG, although no function has yet been assigned to the encoded GTs at present. Of the other floral network genes, GATL5 is known to be involved in RGI formation [38], which may extend to GATL3; CSLD5 has been attributed to cellulose synthesis, although there is a report suggesting a role in mannan biosynthesis [39–41]; GSL12 is a callose synthase [42]; and GALT31a (and possibly AT5G41460) assembles AGPs [43].

Visualization of GT mRNAs at the Wild-Type Shoot Apex Reveals Subdomain-Specific Expression of Cell Wall GTs

Co-expression of different GT family members suggests that the encoded enzymes, and their respective polysaccharides, may contribute to either specific types of walls or developmental stages. We therefore performed systematic RNA in situ hybridization assays to examine the expression patterns of the SAM-expressed GTs at single-cell spatial resolution, including those transcripts that show either relative increases, reductions, or no change between SAM and flower samples. Transcript localization data confirmed that most of the candidate GT-encoded genes exhibit high expression in WT SAM as well as flower primordia. Patterns of expression were found to be similar between WT and *clv3* (Figure S3). Based on the expression pattern, these genes were divided into five categories (Figure 3A; Data S1): type 1 represents a uniform distribution across the apex (e.g., *GSL1*); type 2 represents apical patchy distribution, suggesting flower primordia-specific enrichment (e.g., *At3g14960*); type 3 has intense scattered spots (e.g., *CSLD5*); type 4 represents both spotted and general apical enrichment (e.g., *FUT3*); and type 5 consists of patterns not classified in the above. Within the type 5 group, *GATL6* mRNA gives a high signal in the approximate location of vascular initials and in developing outer whorls of the flower, further confirmed by visualizing the protein (Figures S4A–S4C). A summary of the classification of mRNA localization patterns for GT families, together with representative images of all SAM-expressed cell wall GT mRNA hybridizations, is shown in Figure 3B and Data S1. Some of these patterns are similar to those reported for non-GT-encoding genes in the tomato SAM [44], with *Arabidopsis* types 3 and 4 exhibiting similarities to tomato pattern 5, represented by *histone 2A*, and suggesting a degree of cell division-linked expression [45]. The majority (74/115) of the *Arabidopsis* GT mRNA patterns are classified as type 4. There are 14 genes that have a type 1 pattern. The expression of these genes could also be detected in other tissues (e.g., shoot tissue), suggesting that they might play fundamental roles in building the cell wall. Indeed, genes grouped into this type include *CESA3* for primary wall cellulose and *CSLC4*, *XXT1*, and *FUT1*, which encode XG GTs. This broad expression of GTs is consistent with the broad labeling of the polysaccharides. In contrast to the type 1 *CESA3* pattern, RNAs for two of the subunits of the third polypeptide of the CSC, encoded by either *CESA5* or *CESA6*, gave a type 2 distribution pattern, confirming upregulation in flower primordia. For XG, other GTs include *MUR3* (type 4), *At5g62220* (type 4), and *XXT2* (type 2). The RNA for a putative GalT, encoded by *GT15*, although found at low levels across the SAM, appears to be confined to the inner whorls of the developing flower (Data S1).

For acidic pectic polysaccharide synthesis, the majority of GAUT family transcripts (*GAUT1*, *3*, *4*, *6*, *7*, *9–11*, and *13–15*) have similar expression patterns, in this case a type 4 distribution. In contrast, of the five genes that comprise the GT92 family-encoding enzymes that specifically make β-1,4-galactan [46], the three shoot apex-expressed genes *GALS1*, *GALS2*, and *GALS3* exhibited divergent expression patterns, grouped into types 2, 3, and 4, respectively (Table S2; Data S1). The mRNA pattern of *GALS2* and *GALS3*, concentrated at sites of primordial initiation and during floral development (Figure S3A), was largely consistent with the signal observed with the corresponding epitope (Figure 2I).

Some members of the GT31 and GT29 families have been proposed to be involved in the assembly of the β-1,6-Gal residues that provide the core substitutions of the β-1,3-galactan backbone of type II AGs [43, 47]. Of particular interest is *At1g32930* (*GALT31a*), because (1) its mRNA is florally enriched (Table S2), (2) its patchy expression pattern matches that expected of early-forming flower primordia (Figure S4D; Data S1), and (3) insertion mutants have been found to arrest during embryo development [43]. To confirm whether the patchy expression is indeed confined to regions of organ initiation, we observed a fluorescent reporter of *GALT31a* promoter activity, together with a YFP plasma membrane reporter for observation of cell boundaries. A top view of a confocal z projection shows that the foci of *GALT31a* expression coincide with flower primordia and at sites where primordia would be expected to initiate based on a spiral phyllotactic pattern (Figures S4E and S4F), confirming that, at least for some genes, the patchy mRNA localization coincides with new flower development.

All transcripts encoding known xylan biosynthetic enzymes have a type 4 distribution and, for heteromannan backbone synthesis, *CSLA* expression patterns were found to be either of type 2 or 4 patterns. The Gal substitutions along the backbone appear to be formed through the action of MUCI10/GTL6. *MUCI10* mRNA signal was very high in some cells (Data S1), consistent with the relatively high levels of mRNA found in the transcriptome analysis of the SAM. Callose is made by GT48 (GSL/CALS) family members, and transcripts were localized in either a type 1 (*GSL1*, *5*, *6*, and *10*), type 2 (*GSL8* and *11*), or type 4 (*GSL3* and *12*) pattern.

Expression of GT Genes during Cell Division

We found that the mRNA of the GT2 member *CSLD5* exhibits intense spots (type 3 pattern) representing single cells against a very low background (Figure 3A). Similarly intense spots were also observed for *GALS2* (Figure S3; Data S1). To test whether the spotted pattern represents actively dividing cells, we carried out dual labeling of *CSLD5* and *GALS2* mRNA together with M phase marker *CyclinB1;1* mRNA by using probes labeled with either cyanine 5 or fluorescein. *CSLD5*

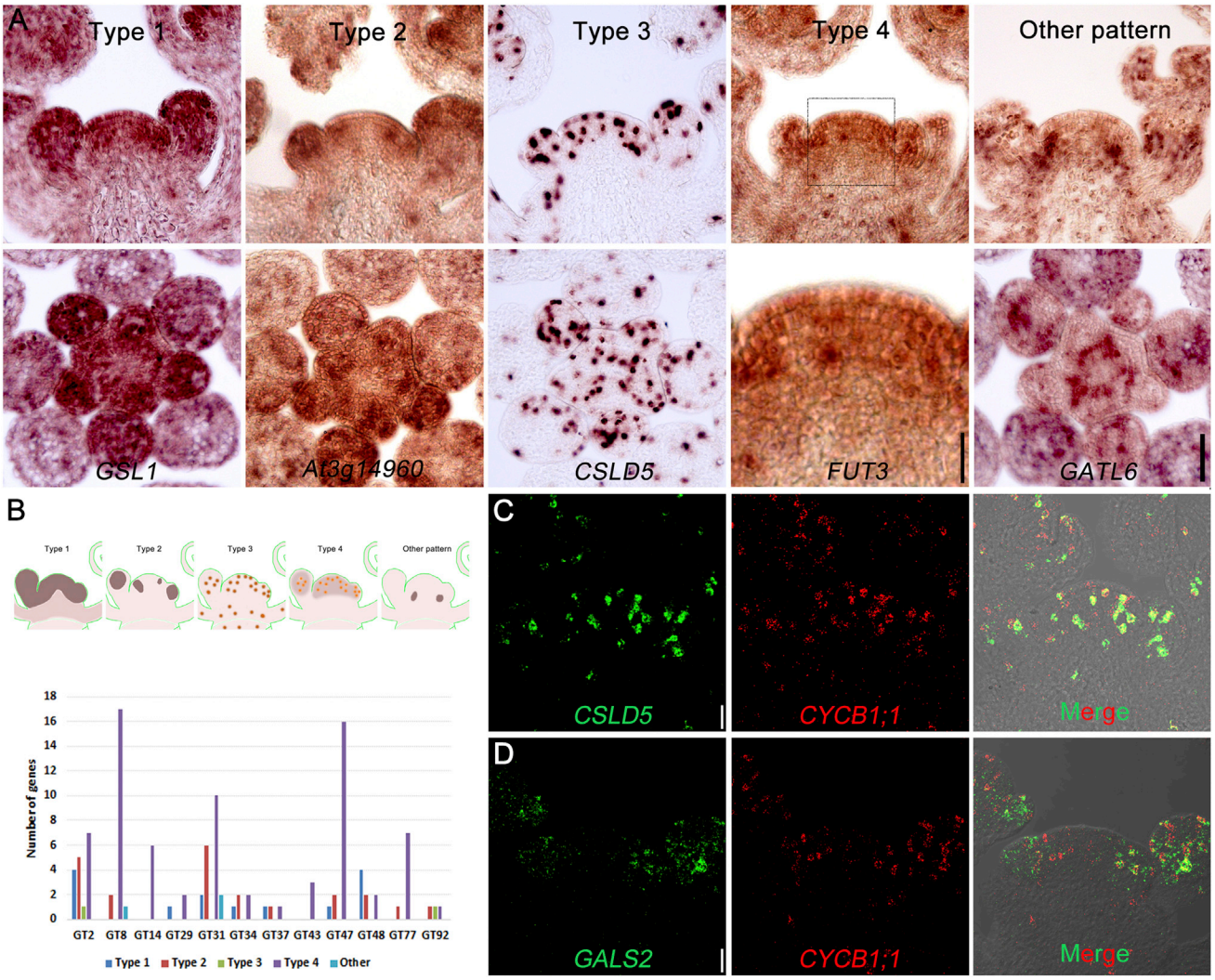


Figure 3. Localization of Glycosyltransferase Gene mRNAs in the Shoot Apex by In Situ Hybridization

(A) The expression patterns of glycosyltransferase genes expressed in the shoot apex. The GTs were classified into five patterns according to their mRNA distribution. One representative gene for each type is shown in both longitudinal (top panel) and transverse (bottom panel) sections except for *FUT3* which shows a magnified region (bottom panel) of the SAM (top panel, boxed area). Scale bars, 50 μm, except for type 4, bottom, 20 μm.

(B) A sketch showing the expression patterns of GTs and a summary of gene number in each GT family, classified into different expression patterns.

(C and D) Co-expression of GT genes with *CyclinB1;1*, which marks dividing cells by dual-labeling fluorescence in situ hybridization. Scale bars, 20 μm.

See also Figures S2–S4, Tables S2 and S3, and Data S1.

mRNA was found in most of the cells containing *CyclinB1;1* mRNA (Figure 3C; Manders coefficient M[1] 0.295), indicating that its expression was linked to the later portion of the cell cycle. *GALS2* also largely co-expressed with *CyclinB1;1* in flower primordia (Figure 3D; M2 0.115). The enriched expression of *CSLD5* and *GALS2* in mitotic cells, together with a significant proportion of GT mRNAs showing an intense punctate labeling (i.e., type 4 pattern), suggests that the expression of a large number of GT genes is upregulated during cell division, probably contributing to the formation of new cross walls.

CSLD Family Genes Function in Stem Cell Maintenance

Given that *CSLD5* and the other shoot apex-expressed members of the *CSLD* family (*CSLD2* and *CSLD3*) are found at comparatively high levels at the shoot apex (Figures 4A–4D; Table S2), we examined SAMs from double- and triple-mutant combinations where whole-plant phenotypes have suggested developmental defects [40]. We found severe growth retardation of the triple-mutant compared to WT (Figures 4E and 4F). The double mutants *csld2 csld5* and *csld3 csld5* and the triple mutant *csld2 csld3 csld5* showed highly similar phenotypes. In our growth conditions, the mutant plants (double and triple mutants) produce only one to three flowers (Figure 4F); however, where flowers did form, they appeared morphologically normal. For some plants no SAM was readily identifiable, especially in the triple mutant, presumably due to early termination. Confocal imaging revealed a small misshapen SAM in the *csld* mutants that was approximately one-quarter of the diameter of WT SAMs,

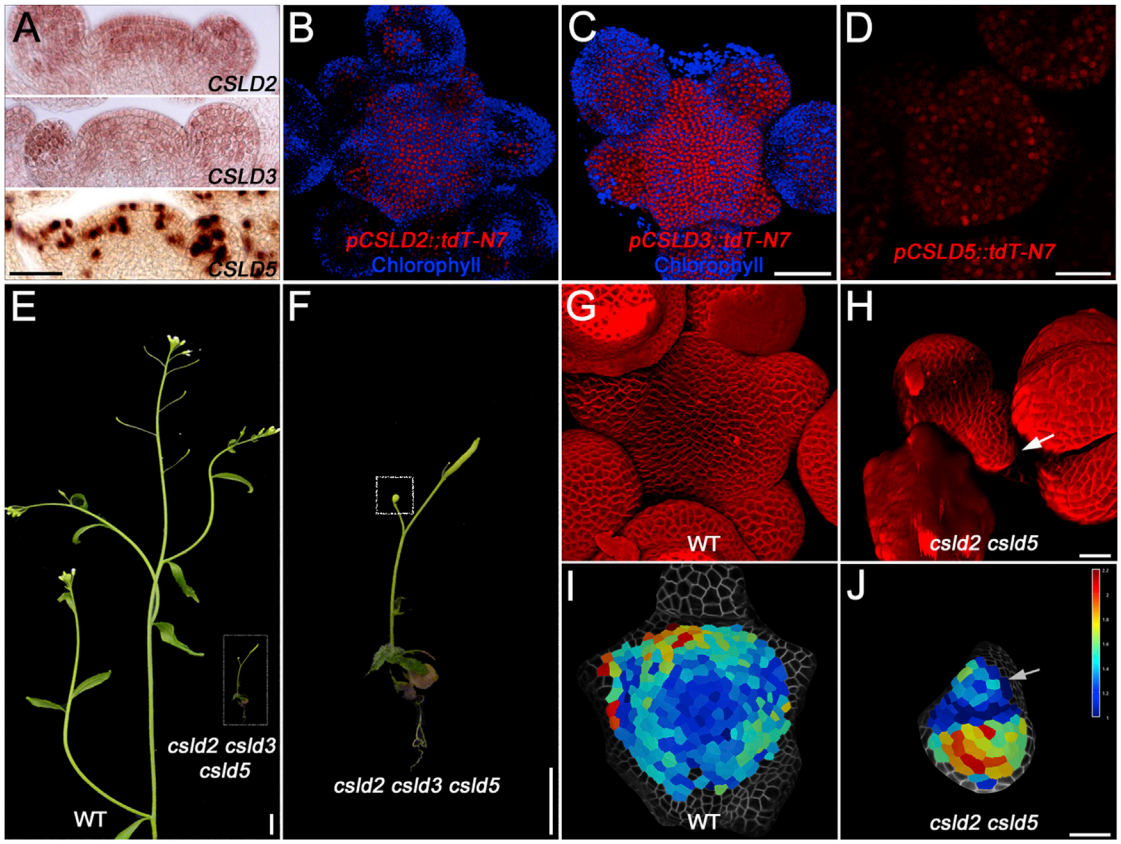


Figure 4. CSLDs Are Required for Shoot Apical Meristem Maintenance

(A) Expression patterns of *CSLD2*, *CSLD3*, and *CSLD5* as revealed by in situ hybridization. Scale bar, 50 μm.

(B–D) Confocal images showing the expression domains of *CSLD2*, *CSLD3*, and *CSLD5* in the SAM. *CSLD2* and *CSLD3* expression is active in most of the cells, whereas *CSLD5* is enriched in dividing cells. Scale bars, 50 μm (C) and 25 μm (D).

(E and F) Whole-plant phenotypes of wild-type (WT, Col-0) and the *csld2 csld3 csld5* triple mutant. A close-up of the *csld2 csld3 csld5* mutant is shown in (F), with the shoot apex boxed. Scale bars, 1 cm (E) and 0.5 mm (F).

(G and H) Three-dimensional rendering of confocal z stacks of wild-type and *csld2 csld5* mutant SAMs. Scale bar, 20 μm.

(I and J) A growth heatmap of wild-type and *csld2 csld5* mutant SAMs showing relative growth increases per cell over a 24-hr period. In the *csld2 csld5* mutant, the location of the SAM is indicated by an arrow (H and J). Scale bar, 20 μm.

See also Figure S5 and Table S3.

with the flowers encompassing a large part of the meristem (compare Figures 4G and 4H). Over a 24-hr period, individual cells were segmented and tracked to compare growth rates. For WT, cells of the CZ yield the lowest growth (blue cells in Figure 4I), with increases observed in the PZ (turquoise through yellow and red in Figure 4I). Cells that are part of flower primordia show the highest degree of growth. For the *csld* mutant, the SAM (Figure 4J, arrow) had no recognizable growth domains; however, the larger flower had cells achieving maximum growth rates consistent with the growth retardation being exerted at the level of the SAM, not the developing flowers. We interpret these data to mean that *CSLD* genes are needed for proper cell growth and proliferation in the SAM, that is, maintaining the meristematic stem cell pool rather than being involved in specifying cell fate.

We also determined the cell wall composition, using linkage analysis, of SAM samples of a *csld3*$^{+/-}$ *csld5* mutant in the *clv3* background. The plants exhibited retarded growth compared to the *clv3* single mutant, and were found to contain cell walls with a marked difference in linkage composition (Figure S5A). The relative abundance of every detectable polysaccharide changed, some dramatically increasing, such as heteromannan (5-fold), and others decreasing, including HG (6-fold), arabinan (2-fold), and XG (3-fold) (Figure S5B).

DISCUSSION

The primary cell wall plays a fundamental role in plant morphogenesis via modulation of cell shape, mechanical feedback, and signaling [5, 10, 12, 13, 48, 49]. Knowing what this wall is made of and how it is made, summarized in Figure 5 [50, 51], allows us to better understand plant shoot development. The SAM uses a reduced set of GTs to make its walls compared to the rest of the plant. Detailed analysis of the transcriptome, and localization of the GT mRNAs revealed that different categories of expression patterns are present across the SAM and young flowers. These expression analyses, together with antibody labeling of wall components and linkage

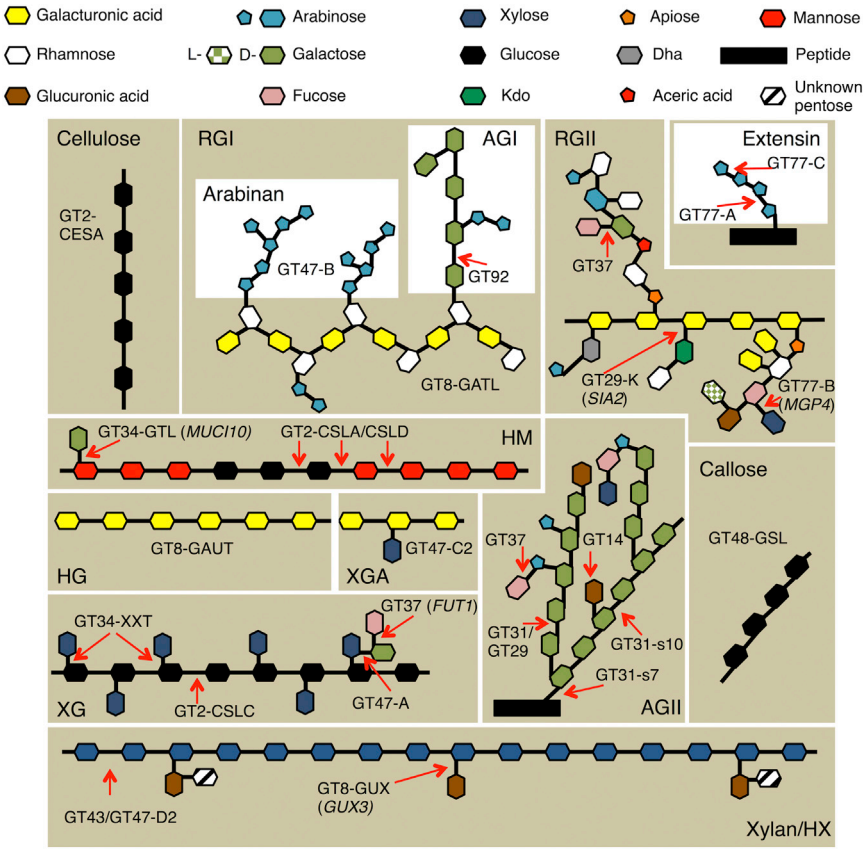


Figure 5. Graphical Representation Showing the Different Types of Polysaccharides and Their Linkages Present in the SAM Cell Wall

Shown are GT family and subfamily assignments. Where only one candidate exists or a transcript is present >10-fold than other members of a GT family, the gene name is given in parentheses. Xylan, RGI, XG, and AG have diverse backbone substitution/branch patterns and the types of linkages are shown. For type II AG(P), the recently identified *Arabidopsis* xylosyl-terminal branch is included [50]. For xylan, the recently determined primary wall structure is shown [51]. The HM shown, as an example, is galactoglucomannan. AGI and AGII, type I and II arabinogalactan; HM, heteromannan; HX, heteroxylan; HG, homogalacturonan; RGI and RGII, type I and II rhamnogalacturonan; XGA, xylogalacturonan; XG, xyloglucan.

analysis, have allowed us to probe the composition of cell walls in the SAM.

Consistent with it being the main structural component of primary cell walls, cellulose is the major polysaccharide in meristem tissues and was found at a reduced level in the SAM compared to the young flower. Whereas the core cellulose subunits, encoded by *CESA1* and *CESA3*, exhibit high expression in both tissues, two of the three genes encoding the third subunit are expressed predominantly in flower primordia.

Pectin is present throughout the primary walls and masks several non-cellulosic polysaccharides, as demonstrated by the requirement for pre-treatment to reveal antibody epitopes. The predominance of pectic arabinan and galactan in specific cell layers or tissues reinforces the idea that both of these polysaccharides are important as part of wall rearrangements during cellular morphogenetic events, likely attributed to their high mobility in the wall and reversible binding to cellulose [52].

For xyloglucan, the action of expansin to release the XG tethers of adjacent cellulose microfibrils has been at the heart of both cellular and tissue-scale models of growth and development, especially in organ emergence from the flanks of the SAM [15]. Consistent with XG being present throughout the apical regions, we found that the only confirmed gene for backbone synthesis, *CSLC4*, showed uniform expression. The dominance of *MUR3* and several GalT genes, as well as the strong and uniform enrichment of *FUT1* mRNA in the SAM, suggests a high degree of galactosylation and fucosylation of XG in meristematic cells (Figure 5).

The xylan synthesis module consists predominantly of GTs encoded by *IRX10L*, *IRX9L* (but not *IRX9*), and *IRX14/14L* for backbone extension and *GUX3* for (methyl)GlcA substitutions, which, based on recent evidence [51], may be extended to include an unidentified pentose moiety (Figure 5). The data presented here support the *IRX-LIKE* genes being associated with primary walls. Given the mostly type 4 distribution of xylan GT transcripts, much of the deposition may occur during the de novo formation of the cross wall during cell division and be maintained as walls mature, which might be involved in the regulation of spacing between cellulose microfibrils [51].

The localization of heteromannan in a small subset of walls that could only be detected after removal of pectin suggests that mannan is either masked by other components and modifications as cell walls mature or principally deposited early during cross-wall formation and later removed. The high expression of *GTL6* in the SAM, recently characterized as the GMGT *MUCI10*, suggests that the heteromannan in the SAM is likely galactosylated to a much higher extent than in other tissues, which might be important for proper cellulose organization [33].

Type II AG represents the glycosidic portion of AGPs that have been implicated in various aspects of plant development [53, 54]. Multiple genes from several AGP-GT families are expressed apically, and their transcripts show type 1, 2, and 4 in situ patterns. *GALT31a* mRNA was found to be particularly abundant in flower primordia, and this was confirmed through live imaging of a reporter of promoter activity. The *atgalt31a* mutant arrests embryo development [43] and may have an equally important role in floral organ formation but, because the mutant is lethal prior to growth of the meristem, functional analysis will require targeted knockdown of the RNA late in development.

In the SAM, active cell division with intervening growth leads to the formation of leaf and flower primordia while also maintaining a pool of stem cells. Compared to animal cells, which divide by forming a constriction, plants build a cell plate at the final step

of cytokinesis to separate the two daughter cells. Callose and cellulose have been implicated as the major components of the cell plate, where callose decorates the nascent cross wall and is gradually replaced by cellulose as the wall matures. Callose is synthesized by members of the *GSL/CALS* GT48 family, exhibiting high expression in the SAM and primordia. No *GSL* transcripts, including *GSL6* [55], were found with the type 3 dotted pattern, suggesting that no GSL enzymes are exclusive to cell-plate formation. Similarly, despite the accumulation of CESA proteins on the forming cell plate [56], none of the *CESA* genes displayed enriched expression in dividing cells, indicating fundamental roles for callose and cellulose in *both* cell-plate formation and wall building during post-mitotic expansion. For callose, this may be limited to surrounding plasmodesmata as part of the regulatory network controlling symplastic transport [42]. Other components of the developing cell plate and their synthesis remain elusive. Antibody labeling suggests the presence of other matrix polysaccharides [57, 58], including mannan (Figure 2M), in the cell plate. Because cell-plate formation is a quite transient process that usually completes in less than 60 min, we can speculate that a large set of GTs would be upregulated through controlled transcription and/or protein synthesis to accommodate a burst of GT enzyme activity for de novo polysaccharide synthesis during cytokinesis. Consistent with this scenario, a majority of GTs exhibited what we interpret as a cell-cycle pattern (type 4) and suggests an important role in transcriptional reprogramming of GT genes upon cell-cycle entry. The decreased expression of these GTs in non-dividing cells also indicated some transient difference in the composition and/or structure between new cross walls and mature walls, and might also imply distinct wall properties of cells within and outside of cell division. However, it is technically challenging to measure the composition of cell plates, due to its transient nature and low abundance in tissues.

The expression of *CSLD5* was strongly linked to cell-division patterns at a time when new cross walls are laid down. *CSLD5* is also found to be co-expressed with markers of the root meristem and division zone [59, 60]. Similar mitosis-enriched expression was observed for the rice homolog *OsCSLD4*, suggesting a conserved regulation of *CSLD* expression across different species [61]. The remaining shoot-expressed members, *CSLD2* and *CSLD3*, displayed broader expression patterns across the apex. *csld2* and *csld3* mutants, when combined with *csld5*, produced plants with terminating meristems, apparently unable to maintain the size of the stem cell pool to produce more than a few flowers. It has been proposed that the *CSLD* gene family encodes GTs that make a type of cellulose [39], with *CSLD5* making a less crystalline polysaccharide [62]; however, there is also a suggestion they may make mannan [40]. A reduction in the *CSLD*-derived polysaccharide resulted in large changes in composition of the SAM cell wall (Figure S5), and demonstrates how changes in wall content and/or wall integrity feed back on wall biosynthesis. Based on our data, we postulate that the polysaccharide made by CSLD is important for proper cell proliferation and cell wall integrity in the SAM, and is not easily compensated by the presence of other wall glycans.

In summary, we find that a limited subset of GTs make the walls in the SAM. Cell wall polysaccharides, and the GTs that make them, can be uniformly distributed or focused to particular regions or cells at a given time. The phenotypes of *csld* mutants demonstrate a clear relationship between the cell wall and the function of the SAM for proper development. The data now allow for a targeted approach for both exploring and manipulating shoot morphogenesis.

EXPERIMENTAL PROCEDURES

Growth and Dissection of *clv3-2* Tissues for RNA-Seq and Cell Wall Linkage Analysis

Because wild-type meristems were too small to obtain sufficient material for linkage analysis (see the Supplemental Experimental Procedures), plants of the *A. thaliana clv3-2* mutant, in the *Ler* background [63], were grown under short-day (8-hr light) regimes for 4 weeks prior to transfer to long days (16 hr). Tissues (two biological replicates of each) were collected from plants exhibiting an inflorescence stem of at least 8 cm. Enlarged SAMs were collected after careful removal of all floral organs. A razor blade was then used to remove the upper fleshy portion at the center of the SAM (Figure 1C, top panel), which was immediately frozen in liquid N_2. Early-stage flowers (stages 6–7) at the periphery of the SAM were collected using fine tweezers and immediately frozen (Figure 1C, bottompanel). Tissues for the *clv3-2* whole-plant sample for RNA-seq (known as the "pooled sample") consisted of fruit (stage 17), young leaves, old leaves, roots, stem, whole inflorescence, and flower (stage 15). The tissues were harvested and stored separately.

RNA-Seq Sample Preparation

RNA was extracted using the RNeasy plant kit (QIAGEN) for each biological replicate according to the manufacturer's protocol. For whole-plant samples, 80 ng of each sample was subsequently pooled before library preparation. Libraries were prepared from 500 ng total RNA using the TruSeq Stranded Total RNA with Ribo-Zero plant kit (Illumina).

qRT-PCR

For qRT-PCR analysis, total RNAs were extracted from dissected meristem, young flowers, and shoot tissues using the RNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. RNA (2 μg) was reverse transcribed into cDNA using oligo(dT) primer and the Transcriptor High Fidelity cDNA synthesis kit (Roche). The cDNA was used as templates for qRT-PCR using LightCycler 480 SYBR Green I Master (Roche) and gene-specific primers (Table S3).

Sequencing Analysis

The six RNA-seq libraries were sent for sequencing at the Beijing Genome Institute using one full lane of Illumina HiSeq 2000. The raw reads in FastQ format were obtained and analyzed in house. We first assessed the quality of the reads using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Potential adaptor contamination and low-quality trailing sequences were removed using Trimmomatic [64]. Preprocessed reads were then mapped to the *Ler*-0 reference genome [65] using TopHat [66, 67], and possible optical duplicates from PCR during the library preparation step were then removed using Picard tools (http://broadinstitute.github.io/picard). Relative transcript estimation was carried out using Cufflinks [67] to obtain fragments per kilobase of transcript per million fragments mapped (FPKM) values and converted to transcripts per million [68], and also as raw reads using HTSeq [69]. Normalization of read counts, gene ontology analysis, and generation of a pairwise correlation coefficient network are described in the Supplemental Experimental Procedures.

Polysaccharide Linkage Analysis

Alcohol-insoluble residues were prepared from isolated SAMs and young flowers and used in the determination of both neutral and acidic monosaccharide linkage composition and polysaccharide composition as previously described [21]. For comparison with *csld*, crosses between *clv3-2* and *csld3 csld5* produced F2 plants where *csld3 csld5 clv3* was not viable beyond early vegetative stage and therefore plants genotyped as *csld3*$^{+/-}$ *csld5 clv3*, which exhibited retarded growth, were used to generate SAM AIRs for linkage analysis.

RNA In Situ Hybridization of Wax-Embedded SAM Sections

For generation of gene-specific probes, cDNA fragments corresponding to each GT gene were amplified using gene-specific primers (Table S3) and ligated into the pGEM-T Easy vector (Promega). The constructs were verified by sequencing and then used as templates for in vitro transcription using the DIG RNA labeling kit (Roche). Shoot apices of *Arabidopsis* wild-type Col-0 or *clv3-9* (a gift from Rüdiger Simon, Heinrich Heine University) were fixed in FAA (formaldehyde, acetic acid, ethanol), embedded in wax, and cut into 8-μm sections. The sections were processed as described (http://www.its.caltech.edu/~plantlab/protocols/insitu.pdf). In brief, after dewaxing, rehydration, and dehydration, the sections were hybridized with gene-specific probes and then incubated with anti-digoxigenin antibody (Roche). Signals were detected by the color reaction after application of NBT/BCIP (nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl-phosphate; Roche). Sense controls, yielding no hybridization with target mRNA, are shown in Figure S3A. Two-color fluorescence in situ hybridization was used for gene co-expression analysis, and the protocol is described, in detail, in the Supplemental Experimental Procedures.

Immunohistochemistry of Wall Components

Wax sections generated for in situ hybridizations were used for antibody detection of wall polymers. Sections were mounted on polished slides and dewaxed using xylene, followed by rehydration using an ethanol series: 100% (v/v), 95%, 70%, 50%, 30%, and 10%. Sections were then incubated in buffer (20 mM Tris-HCl [pH 8.2], 0.5 mM $CaCl_2$, 150 mM NaCl) for 20 min, followed by blocking in 0.5% (w/v) milk powder in buffer. Primary antibody, diluted in buffer, was added to the sections and incubated overnight. After washes in buffer, secondary antibody incubations were carried out for 3 hr. Primary antibodies were rabbit PDM anti-mannan (gift from Paul Dupree, University of Cambridge) [70], LM28 rat anti-xylan (gift from Paul Dupree) [71], LM24 rat anti-xyloglucan [72], JIM7 rat anti-homogalacturonan [73], LM13 rat anti-arabinan [74], and LM5 rat anti-galactan (PlantProbes) [75]. Secondary antibodies were anti-rabbit IgG-CF488A conjugate (Sigma; SAB4600030), anti-rat IgG-CF568 conjugate (Sigma; SAB4600077), and anti-rat Alexa Fluor 647 (Life Technologies). Crystalline cellulose was detected by CBM3a [76] together with anti-His FITC secondary antibody. For enzyme treatments, sections were incubated with pectolyase (0.1% [w/v]; Sigma; P5936) in incubation buffer (0.2 M Na_2HPO_4, 0.1 M citric acid [pH 4.8]) prior to primary antibody incubation. Labeled sections were mounted in ProLong Gold Antifade Mountant (Life Technologies) with a coverslip and sealed. Images were taken with a Zeiss LSM 700 confocal microscope equipped with a 20× 0.8 numerical aperture (NA) dry objective.

Live-Cell Imaging of Transgenic Reporter Lines

Plants were partially dissected to remove overhanging flowers obscuring the SAM. Live-cell imaging was carried out on a Zeiss LSM 700 confocal microscope equipped with 488- and 555-nm lasers and a 20× NA 1.0 water-dipping objective.

Observing Cell Boundaries for Segmentation and Growth Analysis in *csld* Mutants

The *csld2 csld5*, *csld3 csld5*, and *csld2 csld3 csld5* mutants were a gift from Henrik Scheller (University of California, Berkeley) and their construction has been reported [40]. For viewing of the SAM, large organs were dissected and the apex was stained with the dye FM4-64 (Life Technologies; 333 μg ml^{-1}) for 3 min and then carefully rinsed in water. SAMs were viewed under a Zeiss LSM 700 confocal microscope with water-dipping objective, and z stacks were obtained and 3D rendered using confocal software. For growth analyses, the acylated YFP plasma membrane marker (myr-YFP) was transformed into heterozygous plants, and double mutants were identified after selection. Whole plantlets were transferred shortly after bolting to imaging boxes containing 2.2g l^{-1} Murashige and Skoog (MS) medium, Gamborg B5 vitamins (Duchefa), and 1% w/v agar. Confocal imaging of the YFP reporter was then carried out at two time points separated by 24 hr followed by segmentation and cell indexing using MorphoGraphX [77], where a heatmap of relative growth was selected as output upon the 3D image of the 24-hr time point.

ACCESSION NUMBERS

The accession number for the RNA sequencing data reported in this paper is NCBI Sequence Read Archive: SRP072228.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, three tables, and one dataset and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.04.026.

AUTHOR CONTRIBUTIONS

W.Y. and R.W. generated the cell wall preparations and carried out morphological analysis of the *csld* mutants; W.Y. carried out the RNA in situ hybridizations with the help of C.S. and R.W.; C.S. prepared the RNA-seq libraries and carried out the GO analysis; C.S. and V.C. processed the RNA-seq data; C.T.B., A.B., and M.S.D. carried out the linkage analysis and calculated the polysaccharide composition; A.P. devised and performed the cell wall immuno-labeling protocols; R.W. and M.S.D. carried out the literature searches, identified meristem glycosyltransferases, and amalgamated the chemical and expression data; W.Y. and R.W. prepared the figures; and R.W., M.S.D., E.M.M., and A.B. devised the project and wrote the manuscript with contributions from the remaining authors.

ACKNOWLEDGMENTS

The authors wish to thank Zachary Nimchuk for advice on reporter construction, Arun Sampathkumar for help with 3D visualization of the *GALT31A* reporter, and Lisa Willis for help with spherical cap calculations. V.C. is a recipient of a Thailand Research Fund grant for new researchers (grantTRG5880067) and a research supplement grant from the Faculty of Science, Mahidol University. C.T.B., M.S.D., and A.B. acknowledge the support of the ARC Centre of Excellence in Plant Cell Walls (grant CE110001007). E.M.M. acknowledges support from the Gatsby Charitable Trust through fellowships GAT3272/C and GAT3273-PR1, Howard Hughes Medical Institute, Gordon and Betty Moore Foundation (through grant GBMF3406), and U.S. Department of Energy (through award DE-FG02-99ER13873). A.P. acknowledges support of the EU Marie-Curie FP7 COFUND People Programme through the award of an AgreenSkills (grant 267196). R.W. acknowledges support from the Leverhulme Trust (grant RPG-2015-285).

Received: March 11, 2016
Revised: March 24, 2016
Accepted: April 11, 2016
Published: May 19, 2016

REFERENCES

1. Doblin, M.S., Pettolino, F., and Bacic, A. (2010). Plant cell walls: the skeleton of the plant world. Funct. Plant Biol. *37*, 357–381.
2. Cosgrove, D.J. (2005). Growth of the plant cell wall. Nat. Rev. Mol. Cell Biol. *6*, 850–861.
3. Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredez, A., Persson, S., Raab, T., et al. (2004). Toward a systems approach to understanding plant cell walls. Science *306*, 2206–2211.
4. Liepman, A.H., Wightman, R., Geshi, N., Turner, S.R., and Scheller, H.V. (2010). *Arabidopsis*—a powerful model system for plant cell wall research. Plant J. *61*, 1107–1121.
5. Szymanski, D.B., and Cosgrove, D.J. (2009). Dynamic coordination of cytoskeletal and cell wall systems during plant cell morphogenesis. Curr. Biol. *19*, R800–R811.
6. North, H.M., Berger, A., Saez-Aguayo, S., and Ralet, M.-C. (2014). Understanding polysaccharide production and properties using seed coat mutants: future perspectives for the exploitation of natural variants. Ann. Bot. (Lond.) *114*, 1251–1263.

7. Kwiatkowska, D., and Dumais, J. (2003). Growth and morphogenesis at the vegetative shoot apex of *Anagallis arvensis* L. J. Exp. Bot. *54*, 1585–1595.
8. Reinhardt, D., Mandel, T., and Kuhlemeier, C. (2000). Auxin regulates the initiation and radial position of plant lateral organs. Plant Cell *12*, 507–518.
9. Powell, A.E., and Lenhard, M. (2012). Control of organ size in plants. Curr. Biol. *22*, R360–R367.
10. Hamant, O., Heisler, M.G., Jönsson, H., Krupinski, P., Uyttewaal, M., Bokov, P., Corson, F., Sahlin, P., Boudaoud, A., Meyerowitz, E.M., et al. (2008). Developmental patterning by mechanical signals in *Arabidopsis*. Science *322*, 1650–1655.
11. Heisler, M.G., Hamant, O., Krupinski, P., Uyttewaal, M., Ohno, C., Jönsson, H., Traas, J., and Meyerowitz, E.M. (2010). Alignment between PIN1 polarity and microtubule orientation in the shoot apical meristem reveals a tight coupling between morphogenesis and auxin transport. PLoS Biol. *8*, e1000516.
12. Kierzkowski, D., Nakayama, N., Routier-Kierzkowska, A.-L., Weber, A., Bayer, E., Schorderet, M., Reinhardt, D., Kuhlemeier, C., and Smith, R.S. (2012). Elastic domains regulate growth and organogenesis in the plant shoot apical meristem. Science *335*, 1096–1099.
13. Milani, P., Gholamirad, M., Traas, J., Arnéodo, A., Boudaoud, A., Argoul, F., and Hamant, O. (2011). In vivo analysis of local wall stiffness at the shoot apical meristem in *Arabidopsis* using atomic force microscopy. Plant J. *67*, 1116–1123.
14. Peaucelle, A., Braybrook, S.A., Le Guillou, L., Bron, E., Kuhlemeier, C., and Höfte, H. (2011). Pectin-induced changes in cell wall mechanics underlie organ initiation in *Arabidopsis*. Curr. Biol. *21*, 1720–1726.
15. Fleming, A.J., McQueen-Mason, S., Mandel, T., and Kuhlemeier, C. (1997). Induction of leaf primordia by the cell wall protein expansin. Science *276*, 1415–1418.
16. Peaucelle, A., Louvet, R., Johansen, J.N., Höfte, H., Laufs, P., Pelloux, J., and Mouille, G. (2008). *Arabidopsis* phyllotaxis is controlled by the methyl-esterification status of cell-wall pectins. Curr. Biol. *18*, 1943–1948.
17. Verica, J.A., and Medford, J.I. (1997). Modified MERI5 expression alters cell expansion in transgenic *Arabidopsis* plants. Plant Sci. *125*, 201–210.
18. Priestley, J.H. (1929). Cell growth and cell division in the shoot of the flowering plant. New Phytol. *28*, 54–81.
19. Clark, S.E., Running, M.P., and Meyerowitz, E.M. (1995). CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. Development *121*, 2057–2067.
20. Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in *Arabidopsis*. Plant Cell *2*, 755–767.
21. Pettolino, F.A., Walsh, C., Fincher, G.B., and Bacic, A. (2012). Determining the polysaccharide composition of plant cell walls. Nat. Protoc. *7*, 1590–1607.
22. Bromley, J.R., Busse-Wicher, M., Tryfona, T., Mortimer, J.C., Zhang, Z., Brown, D.M., and Dupree, P. (2013). GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. Plant J. *74*, 423–434.
23. Lee, C., Teng, Q., Zhong, R., and Ye, Z.-H. (2012). *Arabidopsis* GUX proteins are glucuronyltransferases responsible for the addition of glucuronic acid side chains onto xylan. Plant Cell Physiol. *53*, 1204–1216.
24. Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z., Segura, M.P., Weimar, T., Yu, X., Seffen, K.A., Stephens, E., Turner, S.R., and Dupree, P. (2010). Absence of branches from xylan in *Arabidopsis* gux mutants reveals potential for simplification of lignocellulosic biomass. Proc. Natl. Acad. Sci. USA *107*, 17409–17414.
25. Desprez, T., Juraniec, M., Crowell, E.F., Jouy, H., Pochylova, Z., Parcy, F., Höfte, H., Gonneau, M., and Vernhettes, S. (2007). Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA *104*, 15572–15577.
26. Sterling, J.D., Atmodjo, M.A., Inwood, S.E., Kumar Kolli, V.S., Quigley, H.F., Hahn, M.G., and Mohnen, D. (2006). Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. Proc. Natl. Acad. Sci. USA *103*, 5236–5241.
27. Atmodjo, M.A., Sakuragi, Y., Zhu, X., Burrell, A.J., Mohanty, S.S., Atwood, J.A., III, Orlando, R., Scheller, H.V., and Mohnen, D. (2011). Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc. Natl. Acad. Sci. USA *108*, 20225–20230.
28. Cocuron, J.-C., Lerouxel, O., Drakakaki, G., Alonso, A.P., Liepman, A.H., Keegstra, K., Raikhel, N., and Wilkerson, C.G. (2007). A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc. Natl. Acad. Sci. USA *104*, 8550–8555.
29. Chou, Y.-H., Pogorelko, G., Young, Z.T., and Zabotina, O.A. (2015). Protein-protein interactions among xyloglucan-synthesizing enzymes and formation of Golgi-localized multiprotein complexes. Plant Cell Physiol. *56*, 255–267.
30. Jensen, J.K., Schultink, A., Keegstra, K., Wilkerson, C.G., and Pauly, M. (2012). RNA-seq analysis of developing nasturtium seeds (*Tropaeolum majus*): identification and characterization of an additional galactosyltransferase involved in xyloglucan biosynthesis. Mol. Plant *5*, 984–992.
31. Vuttipongchaikij, S., Brocklehurst, D., Steele-King, C., Ashford, D.A., Gomez, L.D., and McQueen-Mason, S.J. (2012). *Arabidopsis* GT34 family contains five xyloglucan α-1,6-xylosyltransferases. New Phytol. *195*, 585–595.
32. Edwards, M.E., Dickson, C.A., Chengappa, S., Sidebottom, C., Gidley, M.J., and Reid, J.S.G. (1999). Molecular characterisation of a membrane-bound galactosyltransferase of plant cell wall matrix polysaccharide biosynthesis. Plant J. *19*, 691–697.
33. Voiniciuc, C., Schmidt, M.H.-W., Berger, A., Yang, B., Ebert, B., Scheller, H.V., North, H.M., Usadel, B., and Günl, M. (2015). MUCILAGE-RELATED10 produces galactoglucomannan that maintains pectin and cellulose architecture in *Arabidopsis* seed mucilage. Plant Physiol. *169*, 403–420.
34. Goubet, F., Barton, C.J., Mortimer, J.C., Yu, X., Zhang, Z., Miles, G.P., Richens, J., Liepman, A.H., Seffen, K., and Dupree, P. (2009). Cell wall glucomannan in *Arabidopsis* is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. Plant J. *60*, 527–538.
35. Brown, D.M., Goubet, F., Wong, V.W., Goodacre, R., Stephens, E., Dupree, P., and Turner, S.R. (2007). Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J. *52*, 1154–1168.
36. Brown, D.M., Zhang, Z., Stephens, E., Dupree, P., and Turner, S.R. (2009). Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in *Arabidopsis*. Plant J. *57*, 732–746.
37. De Bodt, S., and Inzé, D. (2013). A guide to CORNET for the construction of coexpression and protein-protein interaction networks. Methods Mol. Biol. *1011*, 327–343.
38. Kong, Y., Zhou, G., Yin, Y., Xu, Y., Pattathil, S., and Hahn, M.G. (2011). Molecular analysis of a family of *Arabidopsis* genes related to galacturonosyltransferases. Plant Physiol. *155*, 1791–1805.
39. Park, S., Szumlanski, A.L., Gu, F., Guo, F., and Nielsen, E. (2011). A role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells. Nat. Cell Biol. *13*, 973–980.
40. Yin, L., Verhertbruggen, Y., Oikawa, A., Manisseri, C., Knierim, B., Prak, L., Jensen, J.K., Knox, J.P., Auer, M., Willats, W.G.T., and Scheller, H.V. (2011). The cooperative activities of CSLD2, CSLD3, and CSLD5 are required for normal *Arabidopsis* development. Mol. Plant *4*, 1024–1037.
41. Wang, W., Wang, L., Chen, C., Xiong, G., Tan, X.-Y., Yang, K.-Z., Wang, Z.-C., Zhou, Y., Ye, D., and Chen, L.-Q. (2011). *Arabidopsis* CSLD1 and CSLD4 are required for cellulose deposition and normal growth of pollen tubes. J. Exp. Bot. *62*, 5161–5177.
42. Vatén, A., Dettmer, J., Wu, S., Stierhof, Y.-D., Miyashima, S., Yadav, S.R., Roberts, C.J., Campilho, A., Bulone, V., Lichtenberger, R., et al. (2011). Callose biosynthesis regulates symplastic trafficking during root development. Dev. Cell *21*, 1144–1155.

43. Geshi, N., Johansen, J.N., Dilokpimol, A., Rolland, A., Belcram, K., Verger, S., Kotake, T., Tsumuraya, Y., Kaneko, S., Tryfona, T., et al. (2013). A galactosyltransferase acting on arabinogalactan protein glycans is essential for embryo development in *Arabidopsis*. Plant J. *76*, 128–137.
44. Fleming, A.J., Mandel, T., Roth, I., and Kuhlemeier, C. (1993). The patterns of gene expression in the tomato shoot apical meristem. Plant Cell *5*, 297–309.
45. Koning, A.J., Tanimoto, E.Y., Kiehne, K., Rost, T., and Comai, L. (1991). Cell-specific expression of plant histone H2A genes. Plant Cell *3*, 657–665.
46. Liwanag, A.J.M., Ebert, B., Verhertbruggen, Y., Rennie, E.A., Rautengarten, C., Oikawa, A., Andersen, M.C.F., Clausen, M.H., and Scheller, H.V. (2012). Pectin biosynthesis: GALS1 in *Arabidopsis thaliana* is a β-1,4-galactan β-1,4-galactosyltransferase. Plant Cell *24*, 5024–5036.
47. Dilokpimol, A., Poulsen, C.P., Vereb, G., Kaneko, S., Schulz, A., and Geshi, N. (2014). Galactosyltransferases from *Arabidopsis thaliana* in the biosynthesis of type II arabinogalactan: molecular interaction enhances enzyme activity. BMC Plant Biol. *14*, 90.
48. Sampathkumar, A., Krupinski, P., Wightman, R., Milani, P., Berquand, A., Boudaoud, A., Hamant, O., Jönsson, H., and Meyerowitz, E.M. (2014). Subcellular and supracellular mechanical stress prescribes cytoskeleton behavior in *Arabidopsis* cotyledon pavement cells. eLife *3*, e01967.
49. Seifert, G.J., and Blaukopf, C. (2010). Irritable walls: the plant extracellular matrix and signaling. Plant Physiol. *153*, 467–478.
50. Tryfona, T., Theys, T.E., Wagner, T., Stott, K., Keegstra, K., and Dupree, P. (2014). Characterisation of FUT4 and FUT6 α-(1→2)-fucosyltransferases reveals that absence of root arabinogalactan fucosylation increases *Arabidopsis* root growth salt sensitivity. PLoS ONE *9*, e93291.
51. Mortimer, J.C., Faria-Blanc, N., Yu, X., Tryfona, T., Sorieul, M., Ng, Y.Z., Zhang, Z., Stott, K., Anders, N., and Dupree, P. (2015). An unusual xylan in *Arabidopsis* primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. Plant J. *83*, 413–426.
52. Lin, D., Lopez-Sanchez, P., and Gidley, M.J. (2016). Interactions of pectins with cellulose during its synthesis in the absence of calcium. Food Hydrocoll. *52*, 57–68.
53. Majewska-Sawka, A., and Nothnagel, E.A. (2000). The multiple roles of arabinogalactan proteins in plant development. Plant Physiol. *122*, 3–10.
54. Johnson, K.L., Kibble, N.A.J., Bacic, A., and Schultz, C.J. (2011). A fasciclin-like arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, fla1, shows defects in shoot regeneration. PLoS ONE *6*, e25154.
55. Hong, Z., Delauney, A.J., and Verma, D.P. (2001). A cell plate-specific callose synthase and its interaction with phragmoplastin. Plant Cell *13*, 755–768.
56. Miart, F., Desprez, T., Biot, E., Morin, H., Belcram, K., Höfte, H., Gonneau, M., and Vernhettes, S. (2014). Spatio-temporal analysis of cellulose synthesis during cell plate formation in *Arabidopsis*. Plant J. *77*, 71–84.
57. Moore, P.J., and Staehelin, L.A. (1988). Immunogold localization of the cell-wall-matrix polysaccharides rhamnogalacturonan I and xyloglucan during cell expansion and cytokinesis in *Trifolium pratense* L.; implication for secretory pathways. Planta *174*, 433–445.
58. Northcote, D.H., Davey, R., and Lay, J. (1989). Use of antisera to localize callose, xylan and arabinogalactan in the cell-plate, primary and secondary walls of plant cells. Planta *178*, 353–366.
59. Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science *318*, 801–806.
60. Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J. (2005). The Botany Array Resource: e-northerns, expression angling, and promoter analyses. Plant J. *43*, 153–163.
61. Yoshikawa, T., Eiguchi, M., Hibara, K., Ito, J., and Nagato, Y. (2013). Rice slender leaf 1 gene encodes cellulose synthase-like D4 and is specifically expressed in M-phase cells to regulate cell proliferation. J. Exp. Bot. *64*, 2049–2061.
62. Bernal, A.J., Jensen, J.K., Harholt, J., Sørensen, S., Moller, I., Blaukopf, C., Johansen, B., de Lotto, R., Pauly, M., Scheller, H.V., and Willats, W.G. (2007). Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in *Arabidopsis*. Plant J. *52*, 791–802.
63. Bowman, J. (1994). *Arabidopsis*: An Atlas of Morphology and Development (Springer-Verlag).
64. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
65. Gan, X., Stegle, O., Behr, J., Steffen, J.G., Drewe, P., Hildebrand, K.L., Lyngsoe, R., Schultheiss, S.J., Osborne, E.J., Sreedharan, V.T., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature *477*, 419–423.
66. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.
67. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.
68. Pachter, L. (2011). Models for transcript quantification from RNA-seq. arXiv, arXiv:1104.3889.
69. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.
70. Handford, M.G., Baldwin, T.C., Goubet, F., Prime, T.A., Miles, J., Yu, X., and Dupree, P. (2003). Localisation and characterisation of cell wall mannan polysaccharides in *Arabidopsis thaliana*. Planta *218*, 27–36.
71. Cornuault, V., Buffetto, F., Rydahl, M.G., Marcus, S.E., Torode, T.A., Xue, J., Crépeau, M.-J., Faria-Blanc, N., Willats, W.G.T., Dupree, P., et al. (2015). Monoclonal antibodies indicate low-abundance links between heteroxylan and other glycans of plant cell walls. Planta *242*, 1321–1334.
72. Pedersen, H.L., Fangel, J.U., McCleary, B., Ruzanski, C., Rydahl, M.G., Ralet, M.-C., Farkas, V., von Schantz, L., Marcus, S.E., Andersen, M.C.F., et al. (2012). Versatile high resolution oligosaccharide microarrays for plant glycobiology and cell wall research. J. Biol. Chem. *287*, 39429–39438.
73. Knox, J.P., Linstead, P.J., King, J., Cooper, C., and Roberts, K. (1990). Pectin esterification is spatially regulated both within cell walls and between developing tissues of root apices. Planta *181*, 512–521.
74. Verhertbruggen, Y., Marcus, S.E., Haeger, A., Verhoef, R., Schols, H.A., McCleary, B.V., McKee, L., Gilbert, H.J., and Knox, J.P. (2009). Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. Plant J. *59*, 413–425.
75. Jones, L., Seymour, G.B., and Knox, J.P. (1997). Localization of pectic galactan in tomato cell walls using a monoclonal antibody specific to (1[→]4)-[beta]-D-galactan. Plant Physiol. *113*, 1405–1412.
76. Blake, A.W., McCartney, L., Flint, J.E., Bolam, D.N., Boraston, A.B., Gilbert, H.J., and Knox, J.P. (2006). Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. J. Biol. Chem. *281*, 29321–29329.
77. de Reuille, P.B., Robinson, S., and Smith, R.S. (2014). Quantifying cell shape and gene expression in the shoot apical meristem using MorphoGraphX. Methods Mol. Biol. *1080*, 121–134.

Computational and Structural Biotechnology Journal 14 (2016) 223–233

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

MicroRNAs and oncogenic transcriptional regulatory networks controlling metabolic reprogramming in cancers

Pannapa Pinweha [a,1], Khanti Rattanapornsompong [a,1], Varodom Charoensawan [a,b], Sarawut Jitrapakdee [a,*]

[a] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
[b] Integrative Computational BioScience (ICBS) Center, Mahidol University, Nakhon Pathom 73170, Thailand

ARTICLE INFO

Article history:
Received 16 March 2016
Received in revised form 25 May 2016
Accepted 27 May 2016
Available online 04 June 2016

Keywords:
Cancer
Metabolism
MicroRNA
Oncogene
Transcriptional regulation network

ABSTRACT

Altered cellular metabolism is a fundamental adaptation of cancer during rapid proliferation as a result of growth factor overstimulation. We review different pathways involving metabolic alterations in cancers including aerobic glycolysis, pentose phosphate pathway, *de novo* fatty acid synthesis, and serine and glycine metabolism. Although oncoproteins, c-MYC, HIF1α and p53 are the major drivers of this metabolic reprogramming, post-transcriptional regulation by microRNAs (miR) also plays an important role in finely adjusting the requirement of the key metabolic enzymes underlying this metabolic reprogramming. We also combine the literature data on the miRNAs that potentially regulate 40 metabolic enzymes responsible for metabolic reprogramming in cancers, with additional miRs from computational prediction. Our analyses show that: (1) a metabolic enzyme is frequently regulated by multiple miRs, (2) confidence scores from prediction algorithms might be useful to help narrow down functional miR-mRNA interaction, which might be worth further experimental validation. By combining known and predicted interactions of oncogenic transcription factors (TFs) (c-MYC, HIF1α and p53), sterol regulatory element binding protein 1 (SREBP1), 40 metabolic enzymes, and regulatory miRs we have established one of the first reference maps for miRs and oncogenic TFs that regulate metabolic reprogramming in cancers. The combined network shows that glycolytic enzymes are linked to miRs via p53, c-MYC, HIF1α, whereas the genes in serine, glycine and one carbon metabolism are regulated via the c-MYC, as well as other regulatory organization that cannot be observed by investigating individual miRs, TFs, and target genes.

1. Overall metabolic reprograming in cancers

In response to overstimulation of growth factor signaling, cancer cells reprogram their metabolism in order to accommodate a high demand for macromolecules during rapid proliferation [1–4]. The hallmark of the above metabolic reprograming is the shift from oxidative phosphorylation to aerobic glycolysis, known as the "Warburg effect" [5]. This phenomenon provides some advantages to the tumors because aerobic glycolysis allows them to survive under hypoxic conditions, while an acidic environment selects a highly aggressive population of cancers to survive and metastasize to distal tissues or organs [3,6]. Cancers are also highly anabolic because they require lipids, protein and nucleic acids as constituents of the structural components of the newly divided cells [2]. This highly anabolic phenotype is partly attributed to the Warburg effect because inhibition of pyruvate entering into the mitochondria results in the redirection of glycolytic intermediates to the pentose phosphate pathway (PPP), which provides biosynthetic precursors for nucleotides and lipids [4]. Furthermore, mitochondrial metabolism of cancers is also reprogrammed toward cataplerosis where substantial amounts of tricarboxylic acid (TCA) cycle intermediates are used as the biosynthetic precursors of lipids and amino acids [2]. Therefore, it is not surprising to see up-regulate expression of key enzymes that catalyze the above biosynthetic pathways in several types of cancers. Fig. 1 shows the overall metabolic reprogramming pathways in cancers together with the key regulatory enzymes.

Abbreviations: ACC, acetyl-CoA carboxylase; ACL, ATP-citrate lyase; BRCA1, breast cancer type 1 susceptibility protein; c-MYC, V-myc avian myelocytomatosis viral oncogene homolog; FAS, fatty acid synthase; FH, fumarate hydratase; G6PD, glucose-6-phosphate dehydrogenase; GDH, glutamate dehydrogenase; GLS, glutaminase; GLUT, glucose transporter; HK, hexokinase; 2-HG, 2-hydroxyglutarate; HIF1α, hypoxia inducible factor 1α; IDH, isocitrate dehydrogenase; miR/miRNA, LDH, lactate dehydrogenase micro RNA; p53, tumor protein p53; PEP, phosphoenolpyruvate; MCT, monocarboxylic acid transporter; ME, malic enzyme; PEPCK, phosphoenolpyruvate carboxykinase; PFK, phosphofructokinase; PHGDH, phosphoglycerate dehydrogenase; PGK, phosphoglycerate kinase (PGK); PSAT, phosphoserine aminotransferase; PSPH, phosphoserine phosphatase; PKM, muscle-pyruvate kinase; PDH, pyruvate dehydrogenase; PC, pyruvate carboxylase; PDK, pyruvate dehydrogenase kinase; PPP, pentose phosphate pathway; SDH, succinate dehydrogenase; SHMT, serine hydroxymethyl transferase; SREBP1, sterol regulatory element binding protein 1; TCA, tricarboxylic acid; TFs, transcription factors.

* Corresponding author at: Department of Biochemistry, Faculty of Science, Mahidol University, Rama 6 Rd, Rajathewee, Bangkok 10400, Thailand. Tel.: +66 2 201 5458; fax: +66 2 354 7174.
E-mail address: sarawut.jit@mahidol.ac.th (S. Jitrapakdee).

[1] Equal authorship.

http://dx.doi.org/10.1016/j.csbj.2016.05.005

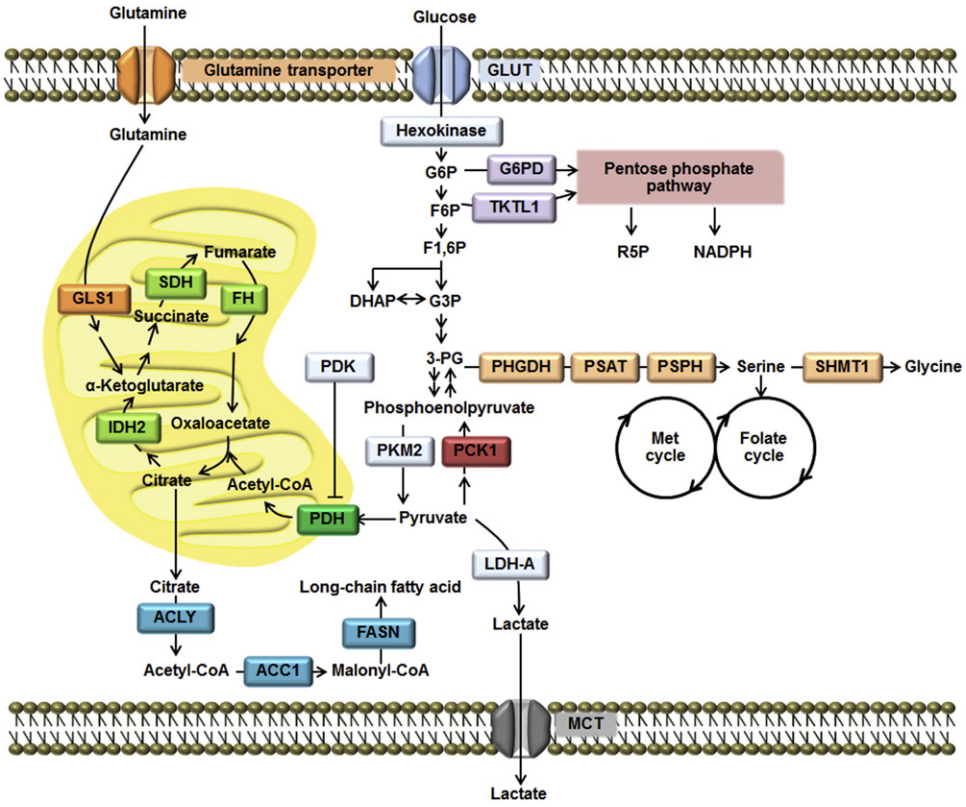


Fig. 1. Metabolic pathways in cancers. Glucose and glutamine are two major carbon sources that are metabolized through these biochemical pathways.

Here we review the altered metabolic pathways and the relevant enzymes in cancers inferred from experimental and computational based data [7–9]. We also review the oncogenic transcription factors (TFs) and miRNAs that regulate those metabolic pathways. In addition, using known and predicted miRNA-target gene interaction, we establish and analyze the network of oncogenic miRNA-metabolic target gene networks that interplay and regulate metabolic reprograming in cancers.

1.1. miRNAs regulate metabolic pathways

Post-transcriptional regulation by microRNAs (miRNAs) has long been known as a mechanism to silence gene expression. miRNAs are short double stranded RNAs, comprising 15–25 nucleotides. They are first transcribed in the nucleus as the primary miRNAs, consisting of multiple stem loop structures, which are then subsequently digested to precursor miRNAs (pre-miRNAs) by Drosha, an RNase III family enzyme [10]. Pre-miRNAs are then transported to the cytoplasm where the hairpin structure is further removed by a dicer enzyme, yielding approximately 21 base pairs miRNA duplex. The miRNA duplex is subsequently incorporated in the Argonaute protein which digests one strand of the duplex miRNA, generating a single stranded miRNA. This single stranded miRNA is further brought to their target mRNAs by an RNA-induced silencing (RISC) complex. Binding of single stranded miRNAs to their targets is mediated by hybridization of 7–8 nucleotides of the miRNAs (known as seed match) to their complementary nucleotides in the 3′-untranslated regions of their targets. Such hybridization results in translational inhibition or degradation of target mRNAs, thus providing a means to inhibit gene expression. Furthermore, one miRNA can bind to more than one species of mRNA targets due to a non-stringent hybridization of the seed match region, allowing simultaneous down-regulation of multiple target mRNAs. In the same way, multiple species of miRNAs can bind to the same mRNA targets and enhance translational inhibition [11]. It is estimated that 45,000 miRNA target sites are found in the human genome, and these miRNAs control expression of up to 60% of human genes [12].

miRNAs are implicated in the regulation of various biological processes. Biochemically, miRNAs also regulate cellular metabolism either directly by targeting key enzymes of metabolic pathways or indirectly by modulating the expression of important transcription factors. Multiple studies have revealed that the altered metabolic pathways in cancers are tightly regulated by miRNAs [13]. In the first half of the review, we describe the metabolic pathways and key enzymes that are altered in various cancers and regulated by miRNAs. This will be followed by the second half on the regulatory networks between metabolic enzymes, regulatory miRNAs and oncogenic transcription factors.

1.2. Glycolytic and pentose phosphate pathways

The Warburg effect is a primary event of metabolic reprogramming during tumorigenesis. This effect includes induced expression of enzymes such as GLUT1, hexokinase 2 (HK2), phosphofructokinase 2 (PFK2) and pyruvate dehydrogenase kinase 1 (PDK1) [3]. Up-regulation of the expression of the first three targets results in a rapid uptake of glucose and increased glycolytic rate, while increased expression of PDK1 inactivates pyruvate dehydrogenase, restricting the conversion of pyruvate to acetyl-CoA in the mitochondria and thus uncoupling glycolysis from subsequent mitochondrial oxidation. Increased expression of lactate dehydrogenase and monocarboxylic acid transporter 4 (MCT4) further sequesters pyruvate toward lactate production, lowering the pH of the extracellular environment [14]. The muscle-specific pyruvate kinase M (PKM) isoform has also been implicated in metabolic reprogramming in certain cancers [15]. PKM exists in two isoforms, PKM1 and PKM2 that have arisen from alternative splicing of exons 9 and 10 [16]. The activities of these two enzymes are determined by their conformers. PKM1 has a tendency to form tetramers that possess high enzymatic activity while PKM2 shows relatively low activity due to its main conformer being dimers. PKM1 is the

most abundant isoform in skeletal muscle while PKM2 is highly expressed during embryonic development. In many cancers, PKM2 is selectively expressed, resulting in the accumulation of phosphoenolpyruvate, and thus redirecting the flow of glycolytic intermediates toward the pentose phosphate pathway (PPP) [15]. This mechanism provides a great benefit for cancers because PPP provides the ribose-5-phosphate and NADPH required for the synthesis of nucleotides and fatty acids. PKM2 also plays a non-metabolic role in which it can act as a co-activator of TFs including HIF1α, STAT3, Oct4 and β-catinin which regulate expression of certain oncogenes [16,17]. Therefore PKM2 switching can reprogram metabolic pathways and alter the program of gene expression in cancers.

In response to PKM2 activation or by other mechanisms, PPP activity has been reported to be elevated in many cancers [18]. Therefore it is not surprising to see up-regulation of key enzymes in this pathway including glucose-6-phosphate dehydrogenase (G6PD), 6-phosphogluconate dehydrogenase (6-PGD) and transketolase-like enzyme [19–21]. NADPH produced by PPP is also crucial for maintaining the proper glutathione-redox loop that cancers use to counter the reactive oxygen species formed especially during epithelial–mesenchymal-transition (EMT) or anoikis resistance [22,23]. Inhibition of PPP via the use of specific enzyme inhibitors or siRNAs targeted to their corresponding enzymes retards growth and biosynthesis of lipid and nucleotides in many types of cancers [21,24,25].

1.3. Mitochondrial metabolism

The tricarboxylic acid cycle (TCA cycle) provides both catabolic and anabolic functions for living cells. In normal cells, the TCA cycle functions as a central oxidation hub where acetyl-CoA derived from oxidations of glucose, amino acids and fatty acids enters for complete oxidation. However in dividing cells or cancers, the TCA cycle is used as an anabolic hub because its intermediates are used as biosynthetic precursors of amino acids, nucleotides and lipids, in a process known as “cataplerosis” [26]. Mutations of certain TCA cycle enzymes such as isocitrate dehydrogenase (IDH), succinate dehydrogenase (SDH) and fumarate hydratase (FH) can contribute to tumorigenesis [27,28]. In certain cancers especially glioma, mutations of the cytosolic (IDH1) or mitochondrial (IDH2) enzymes create a novel function in which they can further convert α-ketoglutarate to 2-hydroxyglutarate (2-HG) [29]. 2-HG is an oncometabolite because it acts as an inhibitor of α-ketoglutarate-dependent dioxygenase involved in DNA and histone demethylation. Inhibition of such a process can lead to tumorigenesis [2,29]. Similarly, mutations of the genes encoding succinate dehydrogenase (SDH) and fumarate hydratase (FH) result in the accumulation of succinate or fumarate, respectively. These two metabolites are inhibitors of prolyl hydroxylase (PHD), which hydroxylates hypoxia-inducible factor 1α (HIF1α), resulting in its degradation by proteolysis. Therefore elevated levels of both metabolites stabilize HIF1α, activating glycolysis in cancers [27].

Cancers also require the replenishment of TCA cycle intermediates after their removal for biosynthetic purposes. In order to prevent a discontinuity in the supply of biosynthetic precursors, there is a biochemical pathway known as “anaplerosis” which is composed of two main reactions, glutaminolysis [30] and pyruvate carboxylation [31]. Glutaminolysis is the conversion of glutamine to glutamate by glutaminase (GLS) before glutamate is further converted to α-ketoglutarate in the TCA cycle by glutamate dehydrogenase. The second anaplerotic reaction is the carboxylation of pyruvate to oxaloacetate by pyruvate carboxylase (PC). Different cancers use these two different anaplerotic reactions to certain extents, to support biosynthesis by up-regulation of either or both enzymes during tumorigenesis [32–35]. Inhibition of these two enzymes results in impaired growth of cancers accompanied with marked reduction in biosynthesis of lipids, nucleotides and amino acids [33–36]. Recent studies show that a gluconeogenic enzyme, phosphoenolpyruvate carboxykinase (PEPCK) also plays an important role in supporting biosynthesis of tumors [37–39]. PEPCK catalyzes a further conversion of oxaloacetate to phosphoenolpyruvate (PEP). This enzyme occurs in two isoforms: the cytosolic (PEPCK1 or PEPCK-C) and the mitochondrial (PEPCK2 or PEPCK-M) isoforms. Colon cancer, for instance, uses PEPCK1 [39] while non-small cell lung cancer uses PEPCK2 [37,38] to supply PEP to support their growth, respectively. However, PEP formed by both enzymes is not only converted to glucose but also used locally as a biosynthetic precursor of serine and glycine. Furthermore, elevated levels of PEP also drive the flow of the upstream glycolytic intermediate glucose-6-phosphate to enter the PPP for the synthesis of ribose sugar required for nucleotide synthesis [37,39]. Interestingly, this function becomes more obvious when the nutrient that supports the growth of a tumor is shifted from glucose to glutamine [37,39]. This adaptive mechanism enables cancers to grow and survive under glucose-limited conditions.

1.4. Amino acid synthesis

Amino acids serve as not only the building blocks of polypeptides, but also the precursors of nucleotides. As cancers require large amounts of proteins and nucleic acids, it is not surprising that up-regulation of key enzymes involved in biosynthesis of certain amino acids were observed in cancer cells. Serine and glycine are essential for synthesis of nucleotides as deprivations of these two amino acids endogenously or exogenously, retard growth of many cancers [40]. *De novo* synthesis of these two amino acids is started from 3-phosphoglycerate (3-PG), an intermediate in the glycolytic pathway. 3-PG is then converted to serine via a three-step reaction, in which 3-PG is first converted to 3-phosphohydroxypyruvate by phosphoglycerate dehydrogenase (PHGDH). 3-phosphohydroxypyruvate is further converted to serine by another two reactions catalyzed by phosphoserine aminotransferase (PSAT) and phosphoserine phosphatase (PSPH) [40]. As only 10% of 3-PG in the glycolytic pool enters serine and glycine biosynthesis, this seems paradoxical with such a high demand for both amino acids during the rapid proliferation of cancers. However, many cancers cope with this limitation via an aberrant activation of the serine biosynthetic pathway by increasing the copy number of the *PHGDH* gene or up-regulating its mRNA expression, resulting in much a higher rate of serine synthesis [41,42]. Serine is further converted to glycine by the serine hydroxymethyl transferase (SHMT), a folate-dependent pathway [40]. SHMT is comprised of two isoforms, SHMT1 which is expressed in the cytoplasm whereas SHMT2 is expressed in mitochondria. It remains unclear about the functional redundancy of these two isoforms as inhibiting activity of either isoform or suppressing their expression retards growth in different cancer models [43–45]. Nevertheless, both SHMT1 and SHMT2 are associated with the folate cycle, which is involved in one-carbon metabolism including synthesis of methionine and nucleotides, and in histone methylation. Thus, disruption of both SHMT isoforms can potentially perturb these metabolic processes [40].

1.5. Lipid biosynthesis

Fatty acids especially in phospholipids are important components of the plasma membrane. In cancers, fatty acids are mainly synthesized through the *de novo* pathway either from glucose or glutamine via glycolysis or glutaminolysis, respectively. However, the latter pathway plays a more significant role in this process [46]. As mentioned earlier, glutamine enters the TCA cycle via glutamate before being converted to α-ketoglutarate by glutamate dehydrogenase. This glutaminolytic flux increases TCA cycle intermediate pools, enabling citrate to leave the mitochondria to enter the cytosol where it is decarboxylated to oxaloacetate and acetyl-CoA by the ATP-citrate lyase (ACL). It has been reported that ACL expression and activity are elevated in many cancers. Thus, inhibition of its activity impairs lipid synthesis and is accompanied by reduced cell growth and survival [47,48]. The cytosolic acetyl-CoA

then serves as a precursor for long chain acyl-CoA synthesis, which is highly regulated by two enzymes, acetyl-CoA carboxylase 1 (ACC1) and fatty acid synthase (FAS). ACC1 catalyzes the carboxylation of acetyl-CoA to form malonyl-CoA, a building block that donates two carbon units for fatty acid synthesis. ACC1 activity can be modulated by a reversible phosphorylation. Among other kinases, the AMP-activated protein kinase (AMP) can phosphorylate ACC1, transforming it into an inactive form while protein phosphatase 1 dephosphorylates ACC1 back to an active form [49]. The phosphorylated ACC1 is subjected to a second mode of regulation through interaction with a DNA repair protein, BRCA1 which is highly expressed in breast tissue [49]. This interaction sequesters phosphorylated ACC1 from being dephosphorylated thereby blocking fatty acid synthesis [50,51]. A high incidence of the oncogene BRCA1 mutations is associated with breast cancer because these mutations not only result in the loss of BRCA1 function as a DNA repair protein but also perturbs its interaction with phosphorylated ACC1, freeing it to be dephosphorylated and subsequently stimulate lipogenesis in breast tissue [51,52]. ACC1 is one of the anti-cancer drug targets because inhibiting its expression or activity induces apoptosis in many cancers [53–55]. FAS has also been reported to be aberrantly activated in many cancers [56–58]. Like ACC1, inhibition of FAS expression or activity markedly reduces cancer growth [52,59,60].

1.6. Metabolic pathway crosstalk contributing to tumorigenesis

Although the crosstalk of signaling pathway is well implicated in tumorigenesis [61], only a few examples of metabolic pathway crosstalk are reported in certain cancers. As mentioned earlier, accumulation of succinate in cancers bearing mutations of succinate dehydrogenase gene not only results in the inactivation of HIF1α, contributing to Warburg effect but this also promotes tumorigenesis by attenuating the production of glutathione, an important redox protein which functions in detoxifying reactive oxygen species (ROS). Several cancers overproduce ROS in order to enhance PI3K, MAPK and NF-κB signaling pathways that support cellular proliferation [1]. Elevated levels of fumarate are found to react with glutathione to form succinated glutathione thereby reducing the NADP/NADPH-couple regeneration system required to eliminate ROS [62]. Similar reduction of glutathione levels was also observed in glioma bearing IDH1 or IDH2 mutation which accumulates 2-HG, suggesting that this oncometabolite may support ROS formation through attenuating the anti-oxidant system [63]. Warburg effect may also enhance tumorigenesis via conversion of fructose-6-phosphate into hexosamine biosynthetic pathway, yielding O-linked N-acetylglucosamine that can enhance mitogenic signaling pathway [64].

1.7. Coordinate regulation of metabolic reprogramming in cancers by oncogenic transcription factors

Having outlined different pathways and mechanisms of metabolic reprogramming in cancers, an important question remains: what controls this metabolic reprogramming in cancers? Three major TFs, namely c-MYC, hypoxia inducible factor 1α (HIF1α) and p53 are responsible for simultaneous up-regulation of the above key metabolic enzymes [65]. Aberrant expression of c-MYC is observed in more than 50% of cancers and it is one of the most amplified oncogenes. The c-MYC regulates various biological processes including proliferation, apoptosis and metabolic reprogramming [66]. Elevated c-MYC levels in turn bind to its target gene promoters, which contain a canonical E-box (CANNTG) element, resulting in increased mRNA transcripts. In normal situations, c-MYC expression is tightly regulated i.e., its expression is high during cell division but rapidly declines during cell cycle arrest [67]. In situations of metabolic alterations, c-MYC targets expression of genes encoding GLUT1, HK2, PDK1 and GLS1 [65,66,68].

The hypoxia-inducible factor (HIF1α), another key oncogenic TF, is functionally coordinated with c-MYC in controlling metabolic reprogramming in cancers [69]. HIF1α exists into two forms: the non-hydroxylated and the hydroxylated forms. In the presence of oxygen, HIF1α undergoes hydroxylation by prolyl hydroxylase, making it prone to proteolysis. However, when oxygen concentration is low, HIF1α escapes hydroxylation, allowing it to enter to the nucleus where it is hetero-dimerized with HIF1β and binds to the hypoxia-responsive element (HRE) in the promoters of genes whose products are involved in angiogenesis and metabolism [3]. HIF1α's metabolic targets appear to overlap with those of c-MYC, including GLUT1, GLUT3, HK1, HK2, aldolase A, phosphoglycerate kinase (PGK), lactate dehydrogenase (LDH), monocarboxylic acid transporter 4 (MCT4), PDK1 and PKM2 [65,70].

Unlike c-MYC and HIF1α, p53 functions as a tumor suppressor protein. Expression of p53 is highly regulated as its expression is essentially low in unstressed cells whereas it becomes highly expressed under stress conditions such as oxidative damage, nutrient limitations and DNA damage [67]. De-regulation of p53 expression caused by mutations is associated with more than half of all cancers [71]. As a transcription factor, p53 binds to the promoter of other tumor suppressor genes such as those involved in cell cycle arrest, DNA repair, apoptosis and metabolism. In addition, p53 can regulate turnover of many proteins independently of transcription [67]. In regard to its regulatory roles on metabolism, p53 inhibits expression of GLUT1, GLUT3, GLUT4, phosphoglycerate mutase 1 (PGM 1), and thus blocking excessive entry of glucose through glycolytic flux [67,72]. p53 inhibits expression of MCT1 and PDK2 while activates expression of PDH1α subunit of PDH complex thereby coupling glycolysis with oxidative phosphorylation [73]. The p53 also down-regulates biosynthesis by decreasing the activity and abundance of glucose-6-phosphate dehydrogenase (G6PD) [74] and decreasing expression of malic enzymes ME1 and ME2 [67,73]. As these three enzymes provide NADPH for biosynthesis, reducing their expression or activities would favor oxidative rather than biosynthetic pathways. In addition to controlling pathways that provide NADPH, p53 can also regulate *de novo* fatty acid synthesis via down-regulating the expression of the sterol regulatory protein 1c (SREBP1c), which is a key transcriptional factor controlling expression of ACL and FAS genes [73]. Therefore, loss-of-function mutations of p53 in cancers literally shift their metabolic phenotype from an oxidative fate to aerobic glycolysis and anabolism. The p53 protein also targets degradation of PEPCK and G6Pase in non-small cell lung cancer [75,76].

1.8. Expanding the repertoire of miRNA target of the alterative expressed metabolic genes in cancer using computational prediction

It has now become clear that many cellular genes including those encoding metabolic enzymes are regulated by miRNAs [13]. Several studies have identified regulatory miRNAs of the key enzymes responsible for metabolic reprogramming while some miRNAs regulate the expression of oncogenic TFs (e.g. c-MYC, HIF1α and p53), which in turn regulate expression of those metabolic enzymes. Despite an increasing number of studies on regulation of metabolic genes through miRNAs in cancers, it is clear that the list of studies on miRNA-regulated metabolic enzymes in cancers is nowhere close to the completion. Furthermore, it is still not known whether some key metabolic enzymes e.g. HK1, Aldolase, MCT4, SHMT2, ACC1, can be regulated by certain miRNAs. Thus, here we sought to explore the repertoire of miRNAs that target expression of key enzymes involved in metabolic reprogramming in cancers by combining known interactions from literature (Table 1) and computational prediction (Supplementary Tables S1 and S2). One of the most important challenges of computational prediction of miRNA is the specificity of the prediction algorithms, which are known to give a large number of false positives. To this end, we examined whether the prediction miRNAs are consistent with the functional validation shown in Table 1, and the predicted miRNA-mRNA interactions that would potentially be worth following up experimentally.

The most frequently used algorithms and webtools currently available for miRNA prediction include miRanda–mirSVR [77,78], DIANA-

Table 1
A list of 40 metabolic enzymes that are involved in metabolic reprogramming in cancers.

Enzyme	Full name	Gene	miRNA	References
Aerobic glycolysis, Warburg effect				
GLUT1	Glucose transporter 1	NM_006516	miR-1291 [123]	[124–126]
GLUT2	Glucose transporter 2	NM_000340	N/A	[124]
GLUT3	Glucose transporter 3	NM_006931	miR-195-5p [127], miR-106-5p [90,128]	[124,129,125,126]
GLUT4	Glucose transporter 4	NM_001042	N/A	[124,130,125]
HK1	Hexokinase1	NM_000188	N/A	[3]
HK2	Hexokinase2	NM_000189	miR-143 [131], [132]	[133,3]
Aldolase A	Aldolase A	NM_000034	N/A	[134]
PGAM1	Phosphoglycerate mutase 1	NM_002629	N/A	[135]
PKM2	Pyruvate kinase 2	NM_002654	miR-122, miR-133a, miR-133b,miR-326 [136–138]	[139,140]
LDHA	Lactate dehydrogenase A	NM_005566	miR-21 [141]	[142,143]
MCT1	Monocarboxylate transporter 1	NM_003051	miR-124 [144]	[145]
MCT4	Monocarboxylate transporter 4	NM_004696	N/A	[145,146]
Pentose phosphate pathway				
G6PD	Glucose-6-phosphate dehydrogenase	NM_000402	miR-206, miR-1 [120]	[20]
TKTL1	Transketolase-like1	NM_012253	miR-206, miR-1 [120]	[19]
Gluconeogenesis				
PCK1	Phosphoenolpyruvate carboxykinase 1	NM_002591	N/A	[39]
PCK2	Phosphoenolpyruvate carboxykinase 2	NM_004563	N/A	[38,37]
Tricarboxylic acid (TCA) cycle				
PDK1	Pyruvate dehydrogenase kinase 1	NM_002610	N/A	[147]
PDH	Pyruvate dehydrogenase	NM_003477	miR-26a [148]	[149]
IDH1	Isocitrate dehydrogenase 1	NM_005896	N/A	[28]
IDH2	Isocitrate dehydrogenase 2	NM_002168	miR-183 [150]	[28]
SDH-B	Succinate dehydrogenase complex iron sulfur subunit B	NM_003000	N/A	[27]
SDH-C	Succinate dehydrogenase complex subunit C	NM_003001	N/A	[27]
SDH-D	Succinate dehydrogenase complex subunit D	NM_003002	miR-210 [151]	[27]
FH	Fumarate hydratase	NM_000143	N/A	[27]
ME1	Malic enzyme 1	NM_002395	N/A	[152]
Glutaminolysis				
GLS1	Glutaminase 1	NM_014905	miR-23a, miR-23b [118]	[32]
GLS2	Glutaminase 2	NM_013267	miR-23a, miR-23b [118]	[153,154]
Serine, Glycine and one carbon metabolism				
SHMT2	Serine hydroxymethyltransferase 2	NM_005412	miR-193b [90,155]	[156]
SHMT1	Serine hydroxymethyltransferase 1	NM_004169	miR-198 [157]	[156]
MTHFD2	Methylenetetrahydrofolate dehydrogenase	NM_006636	miR-9 [158]	[156]
MTHFD1L	Methylenetetrahydrofolate dehydrogenase 1-like	NM_015440	miR-9 [158]	[156]
PHGDH	Phosphoglycerate dehydeogenase	NM_006623	N/A	[41]
PSAT1	Phosphoserine aminotransferase 1	NM_021154	miR-340 [159]	[160,161]
PSPH	Phosphoserine phosphatase	NM_004577	N/A	[161]
GNMT	Glycine-N-methyltransferase	NM_018960	N/A	[162]
de novo fatty acid synthesis				
CIC	Citrate carrier	NM_005984	N/A	[163]
ACLY	ATP citrate lyase Y	NM_001096	N/A	[152,164]
ACC1	Acetyl-CoA carboxylase 1	NM_198836	N/A	[152,165]
FASN	Fatty acid synthase	NM_004104	miR-320 [166]	[58,56,57]
SCD	Stearoyl-CoA desaturase	NM_005063	N/A	[152]

Abbreviation: not available (N/A).

microT-CDS [79], TargetScan [80,81], Pictar [82], miRDB [83], and RNA22 [84], which use common features such as seed match and sequence conservation across the species [85]. In brief, the seed match is a perfect pairing between miRNA and the 3′-UTR of mRNA targets, which usually starts at the 5′ end of miRNA at the positions 2 to 8. There are four main classes of canonical seed matches including (1) 6-mer (6 perfect nucleotide matches between miRNA at positions 2 to 7 and mRNA target), (2) 7mer-A1 (perfect match of miRNA at positions 2 to 7 with an A opposite position 1 of mRNA target), (3) 8-mer (perfect seed paring of miRNA at positions 2 to 8 with an A opposite position 1 of mRNA target) [86] and (4) 7mer-8mer (perfect match of miRNA at positions 2 to 8 and mRNA target) [87,88]. However, these different seed matches do not reflect the degrees of gene expression suppression by miRNAs [89].

With an aim to explore other potential miRNAs that may regulate key metabolic enzymes listed in Table 1, we choose two widely-used miRNA prediction tools that utilize different features to predict miRNA of the target mRNAs of interest, TargetScan7.0 and miRanda–mirSVR. The former predicts the miRNAs targeting a given gene based on the seed match and sequence conservation across the species, whilst the latter uses free energy binding between miRNA and mRNA targets, and the site accessibility for miRNA target prophecy [85]. The context++ scores and mirSVR scores were used as the parameters to indicate the confidence of predictions from the TargetScan7.0 and miRanda–mirSVR, respectively. The context++ score is the sum of contribution from 14 features [81], such as site-type, 3′ pairing, the local AU content [89], target site abundance, seed-pairing stability [80]. The mirSVR scores, on the other hand, can also rank the empirical probability of down-regulation using supervised machine learning of mRNA expression changes as a result of specific microRNA transfection [78]. In short, the more negative context++ scores and mirSVR scores from

the predictions reflect the higher "likelihood" that the mRNA is targeted by miRNA, and thus down-regulated gene expression.

As shown in Fig. 2A, TargetScan7.0 predicted that 40 metabolic enzymes shown in Table 1 are regulated by 299 miRNAs (blue circle). Sixteen out of 40 metabolic enzymes were predicted to be regulated by 113 miRNAs. However, only 8 out of these 113 miRNAs have been reported to functionally regulate expression of these enzymes, leaving the other 105 miRNAs (yellow) whose functional verification is yet to be elucidated. We also noted that there are 14 miRNAs (red) that have been experimentally verified to regulate this set of metabolic genes but elude prediction by TargetScan7.0, suggesting a considerable degree of false negatives. TargetScan7.0 also predicted 186 additional miRNAs that are likely to regulate another 24 metabolic enzymes, whose regulatory miRNAs have not been studied. The list of miRNAs that are predicted to regulate theses 40 metabolic enzymes can be found in Supplementary Table S1.

In a similar trend but not identical, miRanda–mirSVR predicted that there are 395 miRNAs that can potentially regulate these metabolic enzymes (Fig. 2B). One hundred and seventy three miRNAs were predicted to regulate 16 metabolic enzymes while the other 222 miRNAs (gray) were predicted to target another 24 metabolic enzymes which are currently unknown to be regulated by any miRNAs. Within those 16 metabolic enzymes regulated by 173 miRNAs, only 14 miRNAs were independently reported to regulate expression of these metabolic enzymes while the functional verifications of the other 159 miRNAs (pink) are yet to be elucidated. Similar to the TargetScan7.0 prediction but with fewer number of false negatives, eight additional miRNAs have been reported to functionally regulate expression of these 16 metabolic enzymes but were not detected by the miRanda–mirSVR prediction.

Due to the issues of sensitivity and specificity of miRNA prediction algorithms mentioned earlier, we generated boxplots of the context++ scores (Fig. 2C) and mirSRV scores (Fig. 2D), in three miRNA groups: (1) experimentally verified miRNAs with prediction, (2) miRNAs predicted for target genes with other verified miRNAs, but their own functions are yet to be validated, and (3) the predicted miRNAs of metabolic enzymes whose functions have not be validated for any miRNA before (as outlined in the Venn diagrams). We did indeed observe a modest trend that the validated miRNAs have lower context++ scores, than predicted miRNAs without validation; however, the number of miRNAs in each group is likely to be too small to give a statistical significant result. Similarly, the same can be said about the scores assigned to mirSVR prediction, indicating that confidence scores from the prediction might be useful as an extra indicator to extract the predicted miRNA that are likely to be "real" functional miRNAs, and would be worth further experimental validation.

1.9. MicroRNAs and oncogenic transcriptional regulatory networks

To observe the overall interplay of oncogenic TFs, metabolic enzymes, and regulatory miRNAs, we combined the experimentally validated (Table 1), the experimentally validated miRNA-target data from miRTarBase [90] and predicted interactions (from the two algorithms as shown in Fig. 2) into a regulatory network of TFs-metabolic enzymes and miRNA-TFs using Cytoscape [91], as shown in Figs. 3 and 4. Fig. 3 focuses on the known miRNAs that regulate expression of metabolic enzymes via controlling the expression of oncogenic TFs, whereas we expand the network to cover both validated and predicted miRNA-mRNA interactions in Fig. 4. The predicted interactions shown here are the overlaps of the two algorithms used: TargetScan7.0 and miRanda–mirSVR, shown as gray dashed edges, whereas the functional verified miRNA-gene targets from the Table 1 and miRTarBase database [90] are shown in black solid lines. The edges' colors (blue, red, green and purple) represent the miRNAs that regulate expression of metabolic enzymes through the expression of oncogenic TFs (HIF1α, c-MYC, p53, SREBP1, respectively), as in Fig. 3. The colors of node genes in Fig. 4 are classified by metabolic pathways: pale blue color for anaerobic glycolytic genes; white for enzymes involved in serine, glycine and one carbon metabolism; orange for GLS; blue-green nodes for enzymes in the TCA cycle; pink nodes for enzymes in the *de novo* fatty acid synthesis; gray nodes for gluconeogenic enzyme, and purple nodes for enzymes in the pentose phosphate pathway.

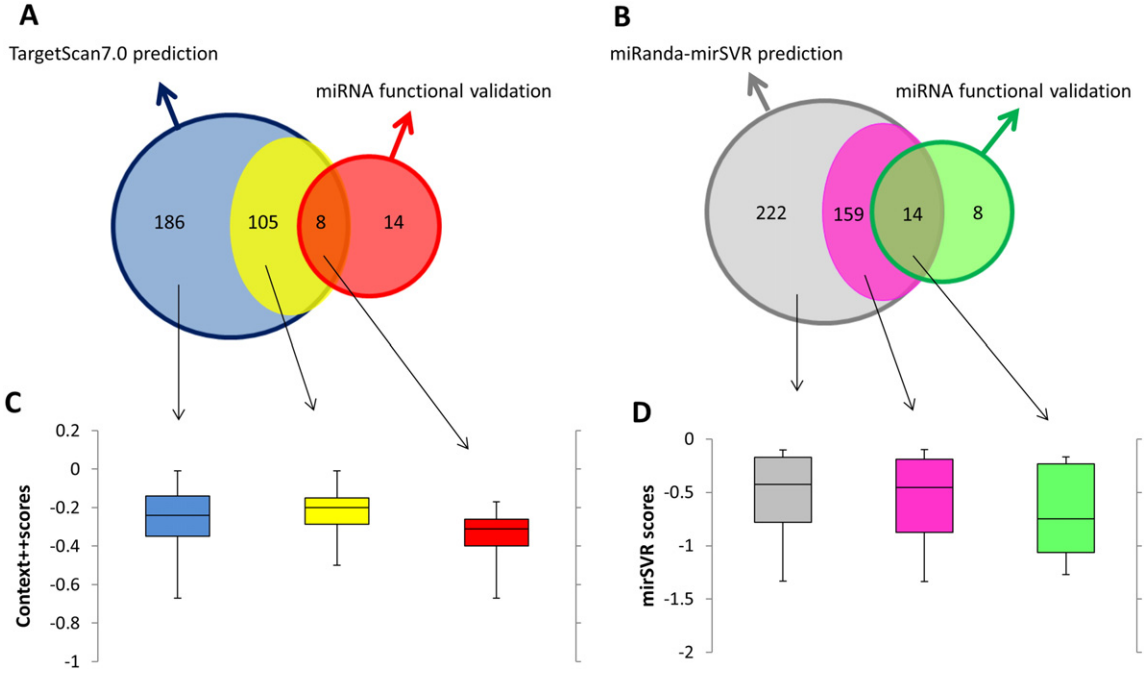


Fig. 2. Venn diagrams and boxplots representing the association between miRNA prediction scores and their functional validation. The Venn diagrams of TargetScan7.0 (Fig. 2A) and miRanda–mirSVR (Fig. 2B) show the numbers of validated and predicted miRNAs that regulate metabolic enzymes in cancers. Boxplots illustrate the association of between context++ scores (Fig. 2C) or miRanda–mirSVR scores (Fig 2D), and three miRNA groups: (1) experimentally validated miRNAs with prediction (2) miRNAs predicted to target metabolic enzymes with other verified miRNAs (3) the predicted miRNAs of altered metabolic enzymes whose functions have not been validated for any miRNA before.

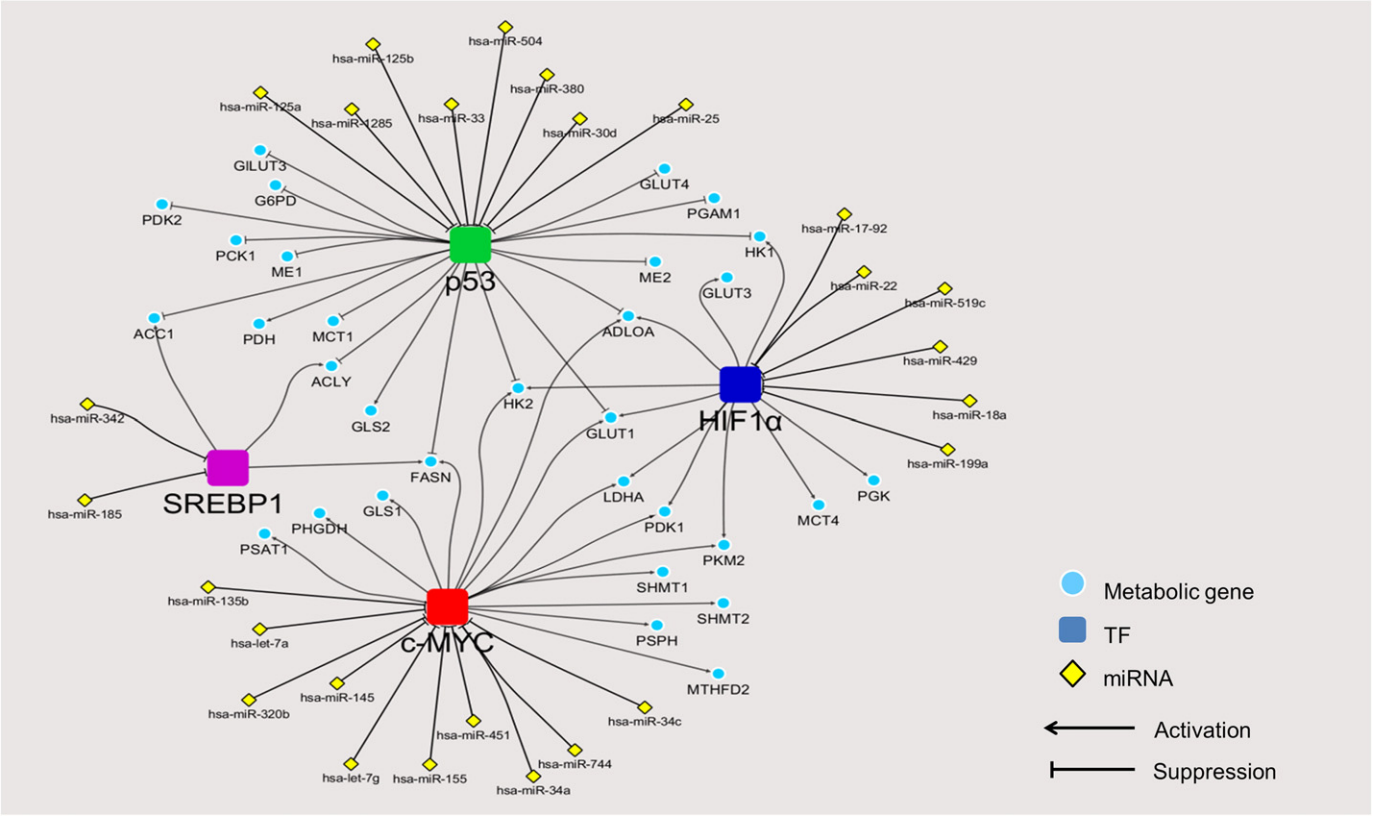


Fig. 3. Regulatory network of experimentally verified miRNAs and oncogenic transcription factors controlling metabolic reprogramming in cancers. The figure shows the integration of experimentally validated regulatory network of TFs-cancer metabolic genes and miRNAs-TFs.

Overall, our miRNAs and oncogenic transcriptional regulatory network depicts individual "modules" of post-transcriptional regulation by miRNA via major drivers of metabolic reprogramming in cancers, acting as hubs that link multiple incoming miRNAs (yellow nodes, Fig. 3) that can bind and suppress transcription of these oncogenes, to their downstream metabolic gene targets (blue nodes). For instance, the expression of c-MYC (red node in Fig. 3, and interaction between miRNA and targeting metabolic genes via c-MYC are in red lines in Fig. 4) is regulated by let-7a in Burkitt Lymphoma [92], miR-145 in non-small cell lung cancer [93], let-7g and miR-744 in hepatocellular carcinoma cells [94,95], miR-34 in prostate cancer cells [96], miR-135b in osteosarcoma cells [97], miR-155 in gastric carcinoma cells [98], miR-320b in colorectal cancer [99] and miR-451 in head and neck squamous cell carcinoma [100]. Suppression of these miRNAs contributes to overexpression of key metabolic enzymes in these tumors. Similarly, HIF1α (dark blue node) expression is regulated by several miRNAs including miR-17-92 in lung cancer cells [101], miR-519c and miR-18a in breast and lung cancer cells [102,103], miR-22 in colon cancer cells [104], miR-199a in non-small cell lung cancers [105] and miR-429 in human endothelial cells [106]. Ectopic expression of these miRNAs reduces the expression of vascular endothelial growth factor (VEGF), a crucial transcriptional target of HIF1α, thereby decreasing angiogenesis, a process of blood vessel formation required for tumor growth and metastasis [107]. Likewise, p53 (green node), a tumor suppressor is also post-transcriptionally regulated by several miRNAs such as miR-25 and miR-30d in myeloma cells [108], miR-125a in breast and hepatoblastoma cells [109], miR-125b in neuroblastoma and lung fibroblalst cells [110], miR-504 in breast and colon cancer cells [111], miR-1285 in neuroblastoma, hepatoblastoma and breast cancer cells [112], miR-33 in hematopoietic stem cells [113] and miR-380 in neuroblastoma cells [114]. Tight regulation of these miRNAs results in substantial expression of p53 which then leads to cell cycle arrest, thus maintaining cells in the non-proliferative state [115]. In contrast, an aberrant overexpression of these p53-target miRNAs results in the down-regulation of p53, causing malignancy. Because this group of miRNAs exerts its effect on the oncogenic transformation, they are generally now classified as the "oncomiR" miRNAs [116].

In addition to these three oncogenes, the sterol regulatory element binding protein (SREBP1, purple node) is also involved in metabolic reprogramming. SREBP1 is a TF that regulates expression of liver type-pyruvate kinase (PKL) and lipogenic enzymes, ACL, ACC and FAS, thus allowing *de novo* fatty acid synthesis from glucose in liver. Cancers also use SREBP1 to up-regulate expression of these lipogenic enzymes to support fatty acid synthesis. Similar to c-MYC, HIF1α and p53, expression of SREBP1 by itself is also regulated by miRNAs. miR-185 and miR-342 play important role in regulation of SREBP1 expression by direct binding to the 3′UTR of its mRNA [117]. Of particular interest, most lipogenic enzymes are co-regulated by more than one TF. For example ACL and ACC1 are regulated by both SREBP1 and p53, while FASN is regulated by SREBP1, p53 and c-MYC. Expression of HK1 is co-regulated by HIF1a and p53 while that of LDHA and PKM2 are co-regulated by HIF1α and c-MYC. GLU1, HK2 and ALDOA are the only three enzymes that are regulated by p53, HIF1α and c-MYC. Interestingly, the expression of certain miRNAs that regulate these metabolic enzymes can also be regulated by an oncogenic TFs. Gao et al. [118] showed that c-MYC indirectly regulates GLS expression in B lymphoma and prostate cancer by suppressing the expression of miR-23a/b that directly regulates the expression of GLS. Kim and coworkers also demonstrated that p53 blocks the expression of HK1, HK2, glucose-6-phosphate isomerase (GPI) and PDK1 by inducing miR-34a expression which in turn, down-regulates the expression of the above four enzymes [119].

Looking at the expanded miRNA–mRNA interaction networks (Fig. 4), we observe a global overview of how metabolic genes involving cancer progression are regulated by miRNA through their direct interaction (black lines for validated interactions and gray lines for those predicted

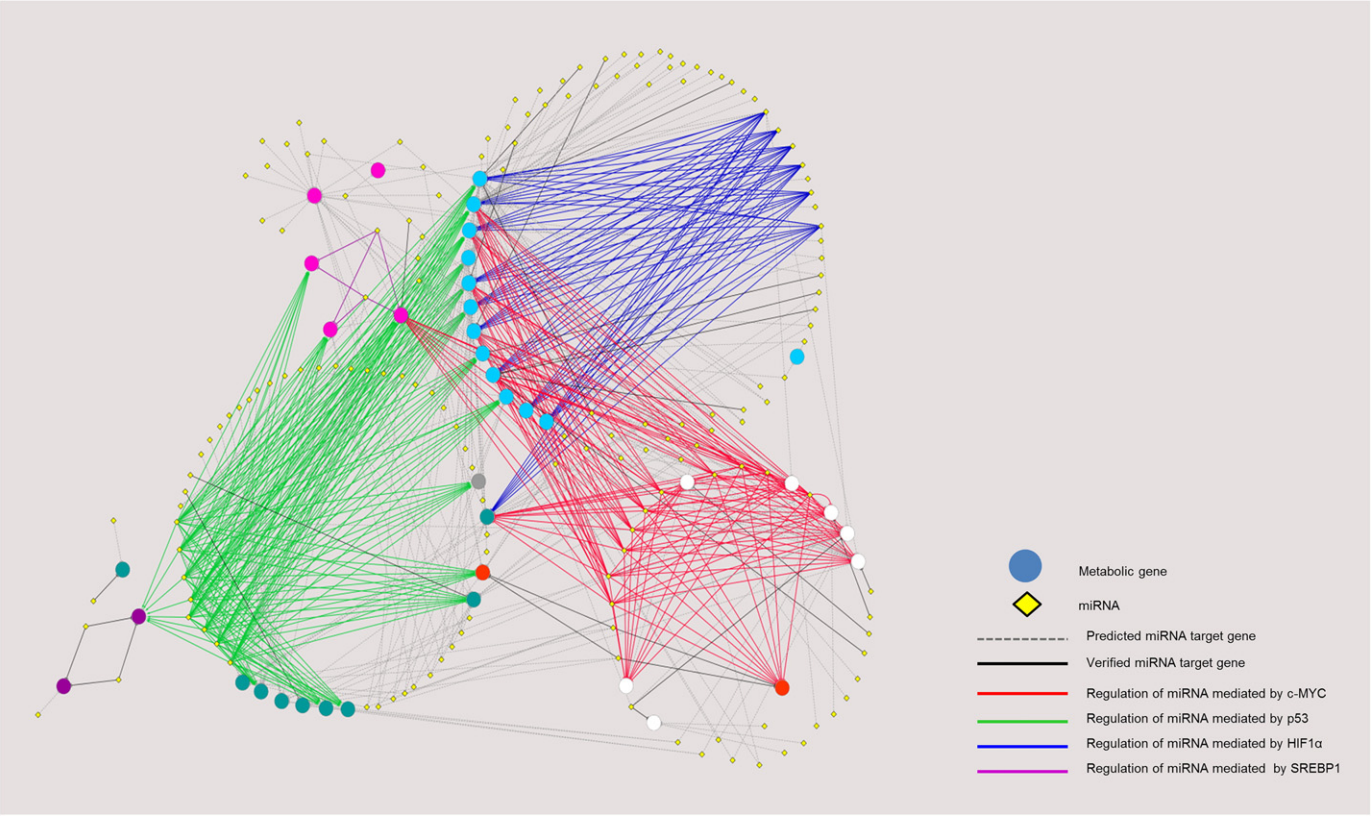


Fig. 4. Regulatory network of miRNAs and oncogenic transcription factors controlling metabolic reprogramming in cancers. The figure shows direct and indirect miRNAs-metabolic genes interaction. The miRNAs that have already verified their regulatory function show in solid edges whereas the dash edges represent the overlap miRNAs from predictions only. In addition, direct interaction of experimentally verified miRNAs and gene targets are showed in black edges whilst the color edges (blue, green, red and purple) illustrate the interaction of miRNAs and cancer metabolic genes via oncogenic transcription factors. Blue edges represent the regulation of miRNA mediated HIF1α, green edges represent the regulation of miRNA mediated p53, red edges represent the regulation of miRNA mediated c-MYC and the purple edges represent the regulation of miRNA mediated SREBP1. The pale blue circle nodes show the anaerobic glycolytic genes, white circle nodes show genes in serine, glycine and one carbon metabolism, orange circle nodes show genes in glutaminolysis, pink circle nodes show genes in *de novo* fatty acid synthesis, purple circle nodes show genes in PPP pathways, gray circle node is PCK1 and the blue-green nodes show genes in TCA cycle. High resolution of the figure with complete labels can be found in Fig. S1.

by TargetScan7.0 and miRanda–mirSVR), or through oncogenic TFs (colored edges). We have seen notable miRNAs such as miR-23a/b that directly control glutaminolysis, whereas the miR-1 and miR-206 are responsible for regulation of the PPP pathway genes, G6PD and TKTL1 [118,120]. The overall network also highlights the "hub" miRNA. miR-429, a tumor suppressor that down-regulates almost all genes in anaerobic glycolytic pathway (e.g. GLUTs) via the oncogenic TF HIF1α. The anaerobic glycolytic genes themselves are also targeted by several other miRNAs such as miR-22, miR-199a, miR-17-92 via HIF1α (blue edges), miR-30d, miR-25, miR-125a/b, miR-1285 via p53 (green edges), and miR-451, miR-155, let-7a, let-7g via c-MYC (red edges). The network also demonstrates other relationships between metabolic pathways and miRNA regulation via TFs. For instance, three out of five genes in *de novo* fatty acid synthesis pathway (ACC1, ACLY, and FASN) share regulation by miRNAs via p53 and SREBP1. The genes in the serine, glycine and one carbon metabolism pathways (white nodes) heavily rely on the regulation of miRNAs via c-MYC. Post-transcriptional regulatory networks have demonstrated intricate regulation of metabolic genes by different miRNAs [13,121,122]. Here, we aim to provide a detailed regulatory network of metabolic genes under direct control of miRNAs, or oncogenic TFs regulated by miRNAs. The high resolution network with complete labels can be found in Supplementary material (Fig. S1 and Table S3). Such overall organization of metabolic gene expression regulation cannot be observed by studying miRNAs, TFs, and target genes individually. Saying that, we note that the current version of network relies on the accuracy of the two prediction algorithms used in this study. The known interactions taken from literature might also be biased toward well-characterized oncogenes such as p53 or c-MYC.

In conclusion, our review not only provides the current status of understanding metabolic reprogramming in cancers but also establishes the regulatory network of miRNA-oncogenic TF-cancer metabolic genes that would provide benefits for research guidance in this emerging field the future.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.csbj.2016.05.005.

Acknowledgements

The authors thank Professor John Wallace, University of Adelaide for helpful comments on the manuscript. PP and KR are supported by the Science Achievement Scholarship of Thailand (SAST), Office of Higher Education Commission. Research in SJ laboratory is supported by grant BRG5780007 from the Thailand Research Fund (TRF) and Mahidol University. VC acknowledges the TRF Grant for New Researcher: TRG5880067, Faculty of Science, Mahidol University, and the Crown Property Bureau Foundation. The authors have no conflict of interest.

References

[1] Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer 2011;11:85–95.

[2] Schulze A, Harris AL. How cancer metabolism is tuned for proliferation and vulnerable to disruption. Nature 2012;491:364–73.
[3] Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles' heel. Cancer Cell 2008;13:472–82.
[4] Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even Warburg did not anticipate. Cancer Cell 2012;21:297–308.
[5] Warburg O. On the origin of cancer cells. Science 1956;123:309–14.
[6] Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 2009;324:1029–33.
[7] Li C, Han J, Yao Q, Zou C, Xu Y, et al. Subpathway-GM: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. Nucleic Acids Res 2013;41, e101.
[8] Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res 2010;38:W71–7.
[9] Li F, Xu Y, Shang D, Yang H, Liu W, et al. MPINet: metabolite pathway identification via coupling of global metabolite network structure and metabolomic profile. Biomed Res Int 2014;2014:325697.
[10] Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 2014;15:509–24.
[11] Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol 2013;14:475–88.
[12] Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009;19:92–105.
[13] Rottiers V, Naar AM. MicroRNAs in metabolism and metabolic disorders. Nat Rev Mol Cell Biol 2012;13:239–50.
[14] Parks SK, Chiche J, Pouyssegur J. Disrupting proton dynamics and energy metabolism for cancer therapy. Nat Rev Cancer 2013;13:611–23.
[15] Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature 2008;452:230–3.
[16] Tamada M, Suematsu M, Saya H. Pyruvate kinase M2: multiple faces for conferring benefits on cancer cells. Clin Cancer Res 2012;18:5554–61.
[17] Luo W, Semenza GL. Emerging roles of PKM2 in cell metabolism and cancer progression. Trends Endocrinol Metab 2012;23:560–6.
[18] Langbein S, Frederiks WM, zur Hausen A, Popa J, Lehmann J, et al. Metastasis is promoted by a bioenergetic switch: new targets for progressive renal cell cancer. Int J Cancer 2008;122:2422–8.
[19] Zhang S, Yue JX, Yang JH, Cai PC, Kong WJ. Overexpression of transketolase protein TKTL1 is associated with occurrence and progression in nasopharyngeal carcinoma: a potential therapeutic target in nasopharyngeal carcinoma. Cancer Biol Ther 2008;7:517–22.
[20] Wang J, Yuan W, Chen Z, Wu S, Chen J, et al. Overexpression of G6PD is associated with poor clinical outcome in gastric cancer. Tumour Biol 2012;33:95–101.
[21] Lin R, Elf S, Shan C, Kang HB, Ji Q, et al. 6-Phosphogluconate dehydrogenase links oxidative PPP, lipogenesis and tumour growth by inhibiting LKB1-AMPK signalling. Nat Cell Biol 2015;17:1484–96.
[22] Radisky DC, Levy DD, Littlepage LE, Liu H, Nelson CM, et al. Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. Nature 2005;436:123–7.
[23] Zhou G, Dada LA, Wu M, Kelly A, Trejo H, et al. Hypoxia-induced alveolar epithelial–mesenchymal transition requires mitochondrial ROS and hypoxia-inducible factor 1. Am J Physiol Lung Cell Mol Physiol 2009;297:L1120–30.
[24] Sukhatme VP, Chan B. Glycolytic cancer cells lacking 6-phosphogluconate dehydrogenase metabolize glucose to induce senescence. FEBS Lett 2012;586:2389–95.
[25] Xu IM, Lai RK, Lin SH, Tse AP, Chiu DK, et al. Transketolase counteracts oxidative stress to drive cancer development. Proc Natl Acad Sci U S A 2016;113:E725–34.
[26] Owen OE, Kalhan SC, Hanson RW. The key role of anaplerosis and cataplerosis for citric acid cycle function. J Biol Chem 2002;277:30409–12.
[27] King A, Selak MA, Gottlieb E. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. Oncogene 2006;25:4675–82.
[28] Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med 2009;360:765–73.
[29] Dang L, White DW, Gross S, Bennett BD, Bittinger MA, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 2009;462:739–44.
[30] Wise DR, Thompson CB. Glutamine addiction: a new therapeutic target in cancer. Trends Biochem Sci 2010;35:427–33.
[31] Jitrapakdee S, Vidal-Puig A, Wallace JC. Anaplerotic roles of pyruvate carboxylase in mammalian tissues. Cell Mol Life Sci 2006;63:843–54.
[32] Huang F, Zhang Q, Ma H, Lv Q, Zhang T. Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. Int J Clin Exp Pathol 2014;7:1093–100.
[33] Sellers K, Fox MP, Bousamra 2nd M, Slone SP, Higashi RM, et al. Pyruvate carboxylase is critical for non-small-cell lung cancer proliferation. J Clin Invest 2015;125:687–98.
[34] Phannasil P, Thuwajit C, Warnnissorn M, Wallace JC, MacDonald MJ, et al. Pyruvate carboxylase is up-regulated in breast cancer and essential to support growth and invasion of MDA-MB-231 cells. PLoS One 2015;10, e0129848.
[35] Cheng T, Sudderth J, Yang C, Mullen AR, Jin ES, et al. Pyruvate carboxylase is required for glutamine-independent growth of tumor cells. Proc Natl Acad Sci U S A 2011;108:8674–9.
[36] Wise DR, DeBerardinis RJ, Mancuso A, Sayed N, Zhang XY, et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proc Natl Acad Sci U S A 2008;105:18782–7.
[37] Vincent EE, Sergushichev A, Griss T, Gingras MC, Samborska B, et al. Mitochondrial phosphoenolpyruvate carboxykinase regulates metabolic adaptation and enables glucose-independent tumor growth. Mol Cell 2015;60:195–207.
[38] Leithner K, Hrzenjak A, Trotzmuller M, Moustafa T, Kofeler HC, et al. PCK2 activation mediates an adaptive response to glucose depletion in lung cancer. Oncogene 2015;34:1044–50.
[39] Montal ED, Dewi R, Bhalla K, Ou L, Hwang BJ, et al. PEPCK coordinates the regulation of central carbon metabolism to promote cancer cell growth. Mol Cell 2015;60:571–83.
[40] Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev Cancer 2013;13:572–83.
[41] Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 2011;476:346–50.
[42] Locasale JW, Grassian AR, Melman T, Lyssiotis CA, Mattaini KR, et al. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nat Genet 2011;43:869–74.
[43] Paone A, Marani M, Fiascarelli A, Rinaldo S, Giardina G, et al. SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation. Cell Death Dis 2014;5, e1525.
[44] Kim D, Fiske BP, Birsoy K, Freinkman E, Kami K, et al. SHMT2 drives glioma cell survival in ischaemia but imposes a dependence on glycine clearance. Nature 2015;520:363–7.
[45] Marani M, Paone A, Fiascarelli A, Macone A, Gargano M, et al. A pyrazolopyran derivative preferentially inhibits the activity of human cytosolic serine hydroxymethyltransferase and induces cell death in lung cancer cells. Oncotarget 2015.
[46] Currie E, Schulze A, Zechner R, Walther TC, Farese Jr RV. Cellular fatty acid metabolism and cancer. Cell Metab 2013;18:153–61.
[47] Bauer DE, Hatzivassiliou G, Zhao F, Andreadis C, Thompson CB. ATP citrate lyase is an important component of cell growth and transformation. Oncogene 2005;24:6314–22.
[48] Hatzivassiliou G, Zhao F, Bauer DE, Andreadis C, Shaw AN, et al. ATP citrate lyase inhibition can suppress tumor cell growth. Cancer Cell 2005;8:311–21.
[49] Tong L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. Cell Mol Life Sci 2005;62:1784–803.
[50] Magnard C, Bachelier R, Vincent A, Jaquinod M, Kieffer S, et al. BRCA1 interacts with acetyl-CoA carboxylase through its tandem of BRCT domains. Oncogene 2002;21:6729–39.
[51] Moreau K, Dizin E, Ray H, Luquain C, Lefai E, et al. BRCA1 affects lipid synthesis through its interaction with acetyl-CoA carboxylase. J Biol Chem 2006;281:3172–81.
[52] Chajes V, Cambot M, Moreau K, Lenoir GM, Joulin V. Acetyl-CoA carboxylase alpha is essential to breast cancer cell survival. Cancer Res 2006;66:5287–94.
[53] Brusselmans K, De Schrijver E, Verhoeven G, Swinnen JV. RNA interference-mediated silencing of the acetyl-CoA-carboxylase-alpha gene induces growth inhibition and apoptosis of prostate cancer cells. Cancer Res 2005;65:6719–25.
[54] Beckers A, Organe S, Timmermans L, Scheys K, Peeters A, et al. Chemical inhibition of acetyl-CoA carboxylase induces growth arrest and cytotoxicity selectively in cancer cells. Cancer Res 2007;67:8180–7.
[55] Zhan Y, Ginanni N, Tota MR, Wu M, Bays NW, et al. Control of cell growth and survival by enzymes of the fatty acid synthesis pathway in HCT-116 colon cancer cells. Clin Cancer Res 2008;14:5735–42.
[56] Rashid A, Pizer ES, Moga M, Milgraum LZ, Zahurak M, et al. Elevated expression of fatty acid synthase and fatty acid synthetic activity in colorectal neoplasia. Am J Pathol 1997;150:201–8.
[57] Pizer ES, Lax SF, Kuhajda FP, Pasternack GR, Kurman RJ. Fatty acid synthase expression in endometrial carcinoma: correlation with cell proliferation and hormone receptors. Cancer 1998;83:528–37.
[58] Kuhajda FP. Fatty acid synthase and cancer: new application of an old pathway. Cancer Res 2006;66:5977–80.
[59] Thupari JN, Pinn ML, Kuhajda FP. Fatty acid synthase inhibition in human breast cancer cells leads to malonyl-CoA-induced inhibition of fatty acid oxidation and cytotoxicity. Biochem Biophys Res Commun 2001;285:217–23.
[60] Knowles LM, Axelrod F, Browne CD, Smith JW. A fatty acid synthase blockade induces tumor cell-cycle arrest by down-regulating Skp2. J Biol Chem 2004;279:30540–5.
[61] Kolch W, Halasz M, Granovskaya M, Kholodenko BN. The dynamic control of signal transduction networks in cancer cells. Nat Rev Cancer 2015;15:515–27.
[62] Sullivan LB, Martinez-Garcia E, Nguyen H, Mullen AR, Dufour E, et al. The proto-oncometabolite fumarate binds glutathione to amplify ROS-dependent signaling. Mol Cell 2013;51:236–48.
[63] Reitman ZJ, Jin G, Karoly ED, Spasojevic I, Yang J, et al. Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. Proc Natl Acad Sci U S A 2011;108:3270–5.
[64] Onodera Y, Nam JM, Bissell MJ. Increased sugar uptake promotes oncogenesis via EPAC/RAP1 and O-GlcNAc pathways. J Clin Invest 2014;124:367–84.
[65] Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. Cell Mol Life Sci 2015;73:377–92.
[66] Dang CV. MYC, metabolism, cell growth, and tumorigenesis. Cold Spring Harb Perspect Med 2013;3.
[67] Kruiswijk F, Labuschagne CF, Vousden KH. p53 in survival, death and metabolic health: a lifeguard with a licence to kill. Nat Rev Mol Cell Biol 2015;16:393–405.
[68] Zeller KI, Jegga AG, Aronow BJ, O'Donnell KA, Dang CV. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. Genome Biol 2003;4:R69.
[69] Dang CV, Kim JW, Gao P, Yustein J. The interplay between MYC and HIF in cancer. Nat Rev Cancer 2008;8:51–6.

[70] Semenza GL. HIF-1: upstream and downstream of cancer metabolism. Curr Opin Genet Dev 2010;20:51–6.
[71] Liu J, Zhang C, Hu W, Feng Z. Tumor suppressor p53 and its mutants in cancer metabolism. Cancer Lett 2015;356:197–203.
[72] Schwartzenberg-Bar-Yoseph F, Armoni M, Karnieli E. The tumor suppressor p53 down-regulates glucose transporters GLUT1 and GLUT4 gene expression. Cancer Res 2004;64:2627–33.
[73] Berkers CR, Maddocks OD, Cheung EC, Mor I, Vousden KH. Metabolic regulation by p53 family members. Cell Metab 2013;18:617–33.
[74] Jiang P, Du W, Wang X, Mancuso A, Gao X, et al. p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. Nat Cell Biol 2011;13:310–6.
[75] Zhang P, Tu B, Wang H, Cao Z, Tang M, et al. Tumor suppressor p53 cooperates with SIRT6 to regulate gluconeogenesis by promoting FoxO1 nuclear exclusion. Proc Natl Acad Sci U S A 2014;111:10684–9.
[76] Goldstein I, Yizhak K, Madar S, Goldfinger N, Ruppin E, et al. p53 promotes the expression of gluconeogenesis-related genes and enhances hepatic glucose production. Cancer Metab 2013;1:9.
[77] Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res 2008;36:D149–53.
[78] Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 2010;11:R90.
[79] Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res 2013;41:W169–73.
[80] Garcia DM, Baek D, Shin C, Bell GW, Grimson A, et al. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol 2011;18:1139–46.
[81] Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015;4.
[82] Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37:495–500.
[83] Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res 2015;43:D146–52.
[84] Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell 2006;126:1203–17.
[85] Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, et al. Common features of microRNA target prediction tools. Front Genet 2014;5:23.
[86] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20.
[87] Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. PLoS Biol 2005;3, e85.
[88] Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003;115:787–98.
[89] Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 2007;27:91–105.
[90] Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res 2016;44:D239–47.
[91] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[92] Sampson VB, Rong NH, Han J, Yang Q, Aris V, et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res 2007;67:9762–70.
[93] Chen Z, Zeng H, Guo Y, Liu P, Pan H, et al. miRNA-145 inhibits non-small cell lung cancer cell proliferation by targeting c-Myc. J Exp Clin Cancer Res 2010;29:151.
[94] Lan FF, Wang H, Chen YC, Chan CY, Ng SS, et al. Hsa-let-7g inhibits proliferation of hepatocellular carcinoma cells by downregulation of c-Myc and upregulation of p16(INK4A). Int J Cancer 2011;128:319–31.
[95] Lin F, Ding R, Zheng S, Xing D, Hong W, et al. Decrease expression of microRNA-744 promotes cell proliferation by targeting c-Myc in human hepatocellular carcinoma. Cancer Cell Int 2014;14:58.
[96] Yamamura S, Saini S, Majid S, Hirata H, Ueno K, et al. MicroRNA-34a modulates c-Myc transcriptional complexes to suppress malignancy in human prostate cancer cells. PLoS One 2012;7, e29722.
[97] Liu Z, Zhang G, Li J, Liu J, Lv P. The tumor-suppressive microRNA-135b targets c-Myc in osteoscarcoma. PLoS One 2014;9, e102621.
[98] Sun S, Sun P, Wang C, Sun T. Downregulation of microRNA-155 accelerates cell growth and invasion by targeting c-Myc in human gastric carcinoma cells. Oncol Rep 2014;32:951–6.
[99] Wang H, Cao F, Li X, Miao H, E J, et al. miR-320b suppresses cell proliferation by targeting c-Myc in human colorectal cancer cells. BMC Cancer 2015;15:748.
[100] Wang H, Zhang G, Wu Z, Lu B, Yuan D, et al. MicoRNA-451 is a novel tumor suppressor via targeting c-Myc in head and neck squamous cell carcinomas. J Cancer Res Ther 2015;11(Suppl. 2):C216–21.
[101] Taguchi A, Yanagisawa K, Tanaka M, Cao K, Matsuyama Y, et al. Identification of hypoxia-inducible factor-1 alpha as a novel target for miR-17-92 microRNA cluster. Cancer Res 2008;68:5540–5.
[102] Cha ST, Chen PS, Johansson G, Chu CY, Wang MY, et al. MicroRNA-519c suppresses hypoxia-inducible factor-1alpha expression and tumor angiogenesis. Cancer Res 2010;70:2675–85.
[103] Krutilina R, Sun W, Sethuraman A, Brown M, Seagroves TN, et al. MicroRNA-18a inhibits hypoxia-inducible factor 1alpha activity and lung metastasis in basal breast cancers. Breast Cancer Res 2014;16:R78.
[104] Yamakuchi M, Yagi S, Ito T, Lowenstein CJ. MicroRNA-22 regulates hypoxia signaling in colon cancer cells. PLoS One 2011;6, e20291.
[105] Ding G, Huang G, Liu HD, Liang HX, Ni YF, et al. MiR-199a suppresses the hypoxia-induced proliferation of non-small cell lung cancer cells through targeting HIF1alpha. Mol Cell Biochem 2013;384:173–80.
[106] Bartoszewska S, Kochan K, Piotrowski A, Kamysz W, Ochocka RJ, et al. The hypoxia-inducible miR-429 regulates hypoxia-inducible factor-1alpha expression in human endothelial cells through a negative feedback loop. FASEB J 2015;29:1467–79.
[107] Deng G, Sui G. Noncoding RNA in oncogenesis: a new era of identifying key players. Int J Mol Sci 2013;14:18319–49.
[108] Kumar M, Lu Z, Takwi AA, Chen W, Callander NS, et al. Negative regulation of the tumor suppressor p53 gene by microRNAs. Oncogene 2011;30:843–53.
[109] Zhang Y, Gao JS, Tang X, Tucker LD, Quesenberry P, et al. MicroRNA 125a and its regulation of the p53 tumor suppressor gene. FEBS Lett 2009;583:3725–30.
[110] Le MT, Teh C, Shyh-Chang N, Xie H, Zhou B, et al. MicroRNA-125b is a novel negative regulator of p53. Genes Dev 2009;23:862–76.
[111] Hu W, Chan CS, Wu R, Zhang C, Sun Y, et al. Negative regulation of tumor suppressor p53 by microRNA miR-504. Mol Cell 2010;38:689–99.
[112] Tian S, Huang S, Wu S, Guo W, Li J, et al. MicroRNA-1285 inhibits the expression of p53 by directly targeting its 3′ untranslated region. Biochem Biophys Res Commun 2010;396:435–9.
[113] Herrera-Merchan A, Cerrato C, Luengo G, Dominguez O, Piris MA, et al. miR-33-mediated downregulation of p53 controls hematopoietic stem cell self-renewal. Cell Cycle 2010;9:3277–85.
[114] Swarbrick A, Woods SL, Shaw A, Balakrishnan A, Phua Y, et al. miR-380-5p represses p53 to control cellular survival and is associated with poor outcome in MYCN-amplified neuroblastoma. Nat Med 2010;16:1134–40.
[115] Hermeking H. MicroRNAs in the p53 network: micromanagement of tumour suppression. Nat Rev Cancer 2012;12:613–26.
[116] Esquela-Kerscher A, Slack FJ. Oncomirs — microRNAs with a role in cancer. Nat Rev Cancer 2006;6:259–69.
[117] Li X, Chen YT, Josson S, Mukhopadhyay NK, Kim J, et al. MicroRNA-185 and 342 inhibit tumorigenicity and induce apoptosis through blockade of the SREBP metabolic pathway in prostate cancer cells. PLoS One 2013;8, e70987.
[118] Gao P, Tchernyshyov I, Chang TC, Lee YS, Kita K, et al. c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature 2009;458:762–5.
[119] Kim HR, Roe JS, Lee JE, Cho EJ, Youn HD. p53 regulates glucose metabolism by miR-34a. Biochem Biophys Res Commun 2013;437:225–31.
[120] Singh A, Happel C, Manna SK, Acquaah-Mensah G, Carrerero J, et al. Transcription factor NRF2 regulates miR-1 and miR-206 to drive tumorigenesis. J Clin Invest 2013;123:2921–34.
[121] Feng L, Xu Y, Zhang Y, Sun Z, Han J, et al. Subpathway-GMir: identifying miRNA-mediated metabolic subpathways by integrating condition-specific genes, microRNAs, and pathway topologies. Oncotarget 2015;6:39151–64.
[122] Vienberg S, Geiger J, Madsen S, Dalgaard LT. MicroRNAs in metabolism. Acta Physiol (Oxf) 2016.
[123] Yamasaki T, Seki N, Yoshino H, Itesako T, Yamada Y, et al. Tumor-suppressive microRNA-1291 directly regulates glucose transporter 1 in renal cell carcinoma. Cancer Sci 2013;104:1411–9.
[124] Noguchi Y, Marat D, Saito A, Yoshikawa T, Doi C, et al. Expression of facilitative glucose transporters in gastric tumors. Hepatogastroenterology 1999;46:2683–9.
[125] Ito T, Noguchi Y, Satoh S, Hayashi H, Inayama Y, et al. Expression of facilitative glucose transporter isoforms in lung carcinomas: its relation to histologic type, differentiation grade, and tumor stage. Mod Pathol 1998;11:437–43.
[126] Kurata T, Oguri T, Isobe T, Ishioka S, Yamakido M. Differential expression of facilitative glucose transporter (GLUT) genes in primary lung cancers and their liver metastases. Jpn J Cancer Res 1999;90:1238–43.
[127] Fei X, Qi M, Wu B, Song Y, Wang Y, et al. MicroRNA-195-5p suppresses glucose uptake and proliferation of human bladder cancer T24 cells by regulating GLUT3 expression. FEBS Lett 2012;586:392–7.
[128] Dai DW, Lu Q, Wang LX, Zhao WY, Cao YQ, et al. Decreased miR-106a inhibits glioma cell glucose uptake and proliferation by targeting SLC2A3 in GBM. BMC Cancer 2013;13:478.
[129] Boado RJ, Black KL, Pardridge WM. Gene expression of GLUT3 and GLUT1 glucose transporters in human brain tumors. Brain Res Mol Brain Res 1994;27:51–7.
[130] Binder C, Binder L, Marx D, Schauer A, Hiddemann W. Deregulated simultaneous expression of multiple glucose transporter isoforms in malignant cells and tissues. Anticancer Res 1997;17:4299–304.
[131] Gregersen LH, Jacobsen A, Frankel LB, Wen J, Krogh A, et al. MicroRNA-143 down-regulates Hexokinase 2 in colon cancer cells. BMC Cancer 2012;12:232.
[132] Jiang S, Zhang LF, Zhang HW, Hu S, Lu MH, et al. A novel miR-155/miR-143 cascade controls glycolysis by regulating hexokinase 2 in breast cancer cells. EMBO J 2012;31:1985–98.
[133] Mathupala SP, Ko YH, Pedersen PL. Hexokinase II: cancer's double-edged sword acting as both facilitator and gatekeeper of malignancy when bound to mitochondria. Oncogene 2006;25:4777–86.
[134] Du S, Guan Z, Hao L, Song Y, Wang L, et al. Fructose-bisphosphate aldolase a is a potential metastasis-associated marker of lung squamous cell carcinoma and promotes lung cell tumorigenesis and migration. PLoS One 2014;9, e85804.

[135] Li C, Shu F, Lei B, Lv D, Zhang S, et al. Expression of PGAM1 in renal clear cell carcinoma and its clinical significance. Int J Clin Exp Pathol 2015;8:9410–5.
[136] Liu AM, Xu Z, Shek FH, Wong KF, Lee NP, et al. miR-122 targets pyruvate kinase M2 and affects metabolism of hepatocellular carcinoma. PLoS One 2014;9, e86872.
[137] Kefas B, Comeau L, Erdle N, Montgomery E, Amos S, et al. Pyruvate kinase M2 is a target of the tumor-suppressive microRNA-326 and regulates the survival of glioma cells. Neuro Oncol 2010;12:1102–12.
[138] Wong TS, Liu XB, Chung-Wai Ho A, Po-Wing Yuen A, Wai-Man Ng R, et al. Identification of pyruvate kinase type M2 as potential oncoprotein in squamous cell carcinoma of tongue through microRNA profiling. Int J Cancer 2008;123:251–7.
[139] Feng C, Gao Y, Wang C, Yu X, Zhang W, et al. Aberrant overexpression of pyruvate kinase M2 is associated with aggressive tumor features and the BRAF mutation in papillary thyroid cancer. J Clin Endocrinol Metab 2013;98:E1524–33.
[140] Zhou CF, Li XB, Sun H, Zhang B, Han YS, et al. Pyruvate kinase type M2 is upregulated in colorectal cancer and promotes proliferation and migration of colon cancer cells. IUBMB Life 2012;64:775–82.
[141] Yang X, Cheng Y, Li P, Tao J, Deng X, et al. A lentiviral sponge for miRNA-21 diminishes aerobic glycolysis in bladder cancer T24 cells via the PTEN/PI3K/AKT/mTOR axis. Tumour Biol 2015;36:383–91.
[142] Xian ZY, Liu JM, Chen QK, Chen HZ, Ye CJ, et al. Inhibition of LDHA suppresses tumor progression in prostate cancer. Tumour Biol 2015;36:8093–100.
[143] Shi M, Cui J, Du J, Wei D, Jia Z, et al. A novel KLF4/LDHA signaling pathway regulates aerobic glycolysis in and progression of pancreatic cancer. Clin Cancer Res 2014;20:4370–80.
[144] Li KK, Pang JC, Ching AK, Wong CK, Kong X, et al. miR-124 is frequently downregulated in medulloblastoma and is a negative regulator of SLC16A1. Hum Pathol 2009;40:1234–43.
[145] Pinheiro C, Longatto-Filho A, Ferreira L, Pereira SM, Etlinger D, et al. Increasing expression of monocarboxylate transporters 1 and 4 along progression to invasive cervical carcinoma. Int J Gynecol Pathol 2008;27:568–74.
[146] Koukourakis MI, Giatromanolaki A, Bougioukas G, Sivridis E. Lung cancer: a comparative study of metabolism related protein expression in cancer cells and tumor associated stroma. Cancer Biol Ther 2007;6:1476–9.
[147] Zabkiewicz J, Pearn L, Hills RK, Morgan RG, Tonks A, et al. The PDK1 master kinase is over-expressed in acute myeloid leukemia and promotes PKC-mediated survival of leukemic blasts. Haematologica 2014;99:858–64.
[148] Chen B, Liu Y, Jin X, Lu W, Liu J, et al. MicroRNA-26a regulates glucose metabolism by direct targeting PDHX in colorectal cancer cells. BMC Cancer 2014;14:443.
[149] Koukourakis MI, Giatromanolaki A, Sivridis E, Gatter KC, Harris AL, et al. Pyruvate dehydrogenase and pyruvate dehydrogenase kinase expression in non-small cell lung cancer and tumor-associated stroma. Neoplasia 2005;7:1–6.
[150] Tanaka H, Sasayama T, Tanaka K, Nakamizo S, Nishihara M, et al. MicroRNA-183 upregulates HIF-1alpha by targeting isocitrate dehydrogenase 2 (IDH2) in glioma cells. J Neurooncol 2013;111:273–83.
[151] Puissegur MP, Mazure NM, Bertero T, Pradelli L, Grosso S, et al. miR-210 is overexpressed in late stages of lung cancer and mediates mitochondrial alterations associated with modulation of HIF-1 activity. Cell Death Differ 2011;18:465–78.
[152] Yahagi N, Shimano H, Hasegawa K, Ohashi K, Matsuzaka T, et al. Co-ordinate activation of lipogenic enzymes in hepatocellular carcinoma. Eur J Cancer 2005;41:1316–22.
[153] Hu W, Zhang C, Wu R, Sun Y, Levine A, et al. Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. Proc Natl Acad Sci U S A 2010;107:7455–60.
[154] Suzuki S, Tanaka T, Poyurovsky MV, Nagano H, Mayama T, et al. Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. Proc Natl Acad Sci U S A 2010;107:7461–6.
[155] Leivonen SK, Rokka A, Ostling P, Kohonen P, Corthals GL, et al. Identification of miR-193b targets in breast cancer cells and systems biological analysis of their functional impact. Mol Cell Proteomics 2011;10(M110):005322.
[156] Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, et al. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science 2012;336:1040–4.
[157] Wu S, Zhang G, Li P, Chen S, Zhang F, et al. miR-198 targets SHMT1 to inhibit cell proliferation and enhance cell apoptosis in lung adenocarcinoma. Tumour Biol 2015;37:5193–202.
[158] Selcuklu SD, Donoghue MT, Rehmet K, de Souza GM, Fort A, et al. MicroRNA-9 inhibition of cell proliferation and identification of novel miR-9 targets by transcriptome profiling in breast cancer cells. J Biol Chem 2012;287:29516–28.
[159] Yan S, Jiang H, Fang S, Yin F, Wang Z, et al. MicroRNA-340 inhibits esophageal cancer cell growth and invasion by targeting phosphoserine aminotransferase 1. Cell Physiol Biochem 2015;37:375–86.
[160] Vie N, Copois V, Bascoul-Mollevi C, Denis V, Bec N, et al. Overexpression of phosphoserine aminotransferase PSAT1 stimulates cell growth and increases chemoresistance of colon cancer cells. Mol Cancer 2008;7:14.
[161] Kim SK, Jung WH, Koo JS. Differential expression of enzymes associated with serine/glycine metabolism in different breast cancer subtypes. PLoS One 2014;9, e101004.
[162] Song YH, Shiota M, Kuroiwa K, Naito S, Oda Y. The important role of glycine N-methyltransferase in the carcinogenesis and progression of prostate cancer. Mod Pathol 2011;24:1272–80.
[163] Catalina-Rodriguez O, Kolukula VK, Tomita Y, Preet A, Palmieri F, et al. The mitochondrial citrate transporter, CIC, is essential for mitochondrial homeostasis. Oncotarget 2012;3:1220–35.
[164] Qian X, Hu J, Zhao J, Chen H. ATP citrate lyase expression is associated with advanced stage and prognosis in gastric adenocarcinoma. Int J Clin Exp Med 2015;8:7855–60.
[165] Wang C, Rajput S, Watabe K, Liao DF, Cao D. Acetyl-CoA carboxylase-a as a novel target for cancer therapy. Front Biosci (Schol Ed) 2010;2:515–26.
[166] Cheng C, Chen ZQ, Shi XT. MicroRNA-320 inhibits osteosarcoma cells proliferation by directly targeting fatty acid synthase. Tumour Biol 2014;35:4177–83.

Jinawath *et al. J Transl Med (2016) 14:324*
DOI 10.1186/s12967-016-1078-3

Journal of Translational Medicine

REVIEW Open Access

Bridging the gap between clinicians and systems biologists: from network biology to translational biomedical research

Natini Jinawath[1,2†], Sacarin Bunbanjerdsuk[2†], Maneerat Chayanupatkul[3,4], Nuttapong Ngamphaiboon[5], Nithi Asavapanumas[6], Jisnuson Svasti[1,7,8] and Varodom Charoensawan[1,7,9*]

Abstract

With the wealth of data accumulated from completely sequenced genomes and other high-throughput experiments, global studies of biological systems, by simultaneously investigating multiple biological entities (e.g. genes, transcripts, proteins), has become a routine. Network representation is frequently used to capture the presence of these molecules as well as their relationship. Network biology has been widely used in molecular biology and genetics, where several network properties have been shown to be functionally important. Here, we discuss how such methodology can be useful to translational biomedical research, where scientists traditionally focus on one or a small set of genes, diseases, and drug candidates at any one time. We first give an overview of network representation frequently used in biology: what nodes and edges represent, and review its application in preclinical research to date. Using cancer as an example, we review how network biology can facilitate system-wide approaches to identify targeted small molecule inhibitors. These types of inhibitors have the potential to be more specific, resulting in high efficacy treatments with less side effects, compared to the conventional treatments such as chemotherapy. Global analysis may provide better insight into the overall picture of human diseases, as well as identify previously overlooked problems, leading to rapid advances in medicine. From the clinicians' point of view, it is necessary to bridge the gap between theoretical network biology and practical biomedical research, in order to improve the diagnosis, prevention, and treatment of the world's major diseases.

Keywords: Network biology, Systems biology, Biomedical research, Cancers, Personalized therapy

Background

Next-generation sequencing (NGS) and other high-throughput experiments highlight one of the most significant advances in molecular biology over the past decade. Such technological improvements enable a large number of molecules, including genes, transcripts, and proteins to be simultaneously measured in different conditions over time. This rapid generation of data has transformed molecular biology from a "data poor" to "data rich" discipline, leading to the emergence of systems biology [1–4]. The key challenges and bottlenecks of the modern-day molecular biology have shifted from simply gathering information to the analysis and interpretation of large quantities of data that can now be obtained.

Network representations have been widely used in physics and social science for decades, and are now among the most frequently used tools in systems biology. This technique provides not only a systematic representation of both the presence and abundance of biological molecules, but also displays the relationships or interactions between them. Networks have been used to represent the interactions between different types of biological molecules, e.g. protein–protein interactions [5–8], and in various biological systems including transcriptional

*Correspondence: varodom.cha@mahidol.ac.th
†Natini Jinawath and Sacarin Bunbanjerdsuk contributed equally to this work
[7] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand
Full list of author information is available at the end of the article

regulation [9–11], signaling [12–14], and metabolic pathways [15, 16]. Analyses of network sub-structures have revealed fundamental insights into how biological molecules are organized [17–20], which would not have been possible by studying individual genes or proteins.

Network representation and analysis has been successfully applied to study many systems in molecular biology [21]; however, the use of these tools in translational medicine and drug discovery is relatively new [22–24]. This might be due in part to the knowledge and understanding gaps between clinicians and systems biologists. By convention, clinicians typically focus on specific sets of key genetic markers associated with diseases, to identify the most probable drug targets. In contrast, systems biologists have strong computational and analytical skills, but frequently lack hands-on experimental experience. The lack of interaction of systems biologists with patients can prevent a full appreciation of the complexity of the problems and hindrances in biomedical research [25, 26]. In this review, we aim to improve the understanding of challenges in biomedical research and establish a common ground between clinicians and systems biologists to further promote the application of network biology in translational medicine.

Network biology in a nutshell

What are networks; what do they represent?

We first outline the fundamental concepts of a network representation. In general, a network represents the presence of objects or entities in a system as "nodes", and the relationships or interactions among the nodes are called "edges" (Fig. 1). In biology, nodes can represent biological molecules such as genes, proteins, and ligands, or even larger entities such as cells or individual humans. Edges represent physical interactions or contacts between biological molecules, biochemical processes between substrates and products, genetic interactions between genes, and in some cases, interactions between cells or individual organisms.

Biological information described in a network is not restricted to the presence of nodes and their relationships. The size of node, for instance, can reflect abundance of biological molecules (e.g. gene expression levels). Nodes can also be drawn in different shapes and/or colors according to the classification of interest (e.g. gene/protein family). Likewise, the thickness of an edge or the distance between nodes may represent the frequency or strength of pairwise interaction (e.g. affinity of protein–protein interaction); whereas colors can indicate different types of interactions (e.g. physical or genetic interaction). In addition, edges can be directional or non-directional, solid or dotted, depending on the types of interactions. Thus, networks are information-rich representations, which are widely used to summarize, visualize, and analyze large-scale datasets obtained from high-throughput experiments. To give an overview of the current application of networks in biomedical-related fields, here we review two major types of biological networks.

Interaction networks

We first illustrate the components of interaction networks, where the edges represent a "direct" relationship between nodes (Fig. 1, left). For instance, protein interaction networks, i.e. interactomes, describe physical interactions between proteins, usually obtained from high-throughput screening techniques such as yeast-two hybrid [6, 27], or affinity purification followed by mass spectrometry [5, 28]. In humans, analyses of protein–protein interaction networks have shown that dysfunctional interactions can lead to several diseases including neurological disorders such as ataxias [29], autism [30], several types of cancers including breast [31] and colorectal cancers [32], acute lymphoblastic leukemia [33], as well as other inheritable genetic diseases [34–37].

Transcriptional regulation networks (also known as Gene Regulatory Networks, GRNs) are widely used to illustrate the binding events of regulatory proteins, such as transcription factors, to the promoters of targeted genes, and this technique has been employed in the analysis of bacteria [38], budding yeasts [9], worms [39], and embryonic stem cells [40, 41]. GRNs are directional, and the relationship between two nodes is represented by an arrow starting from a regulator and pointing toward a targeted gene. Mis-regulation of gene expression leads to various diseases especially cancers, as seen in the genome-wide transcription network of the vertebrate transcription factor SOX4 [42], and the androgen receptor, a transcription factor that regulates the onset and progression of prostate cancer [43].

Interaction networks have also been used to describe the binding and affinity of ligands or small molecules to targeted proteins. As seen in a drug-target network [44], a list of drugs approved by the Food and Drug Administration (FDA) were linked to proteins according to drug-target binary associations. The analysis of these networks revealed that many drugs have overlapping but not identical sets of targets. In addition, the network analysis indicated that new drugs tend to be, at least partly, linked to well-characterized proteins already targeted by previously developed drugs. This suggests that the pharmaceutical industry might be shifting toward polypharmacology, to systematically address complex diseases using multiple drugs aimed at multiple specific targets in related pathways to improve treatment efficacy [45, 46].

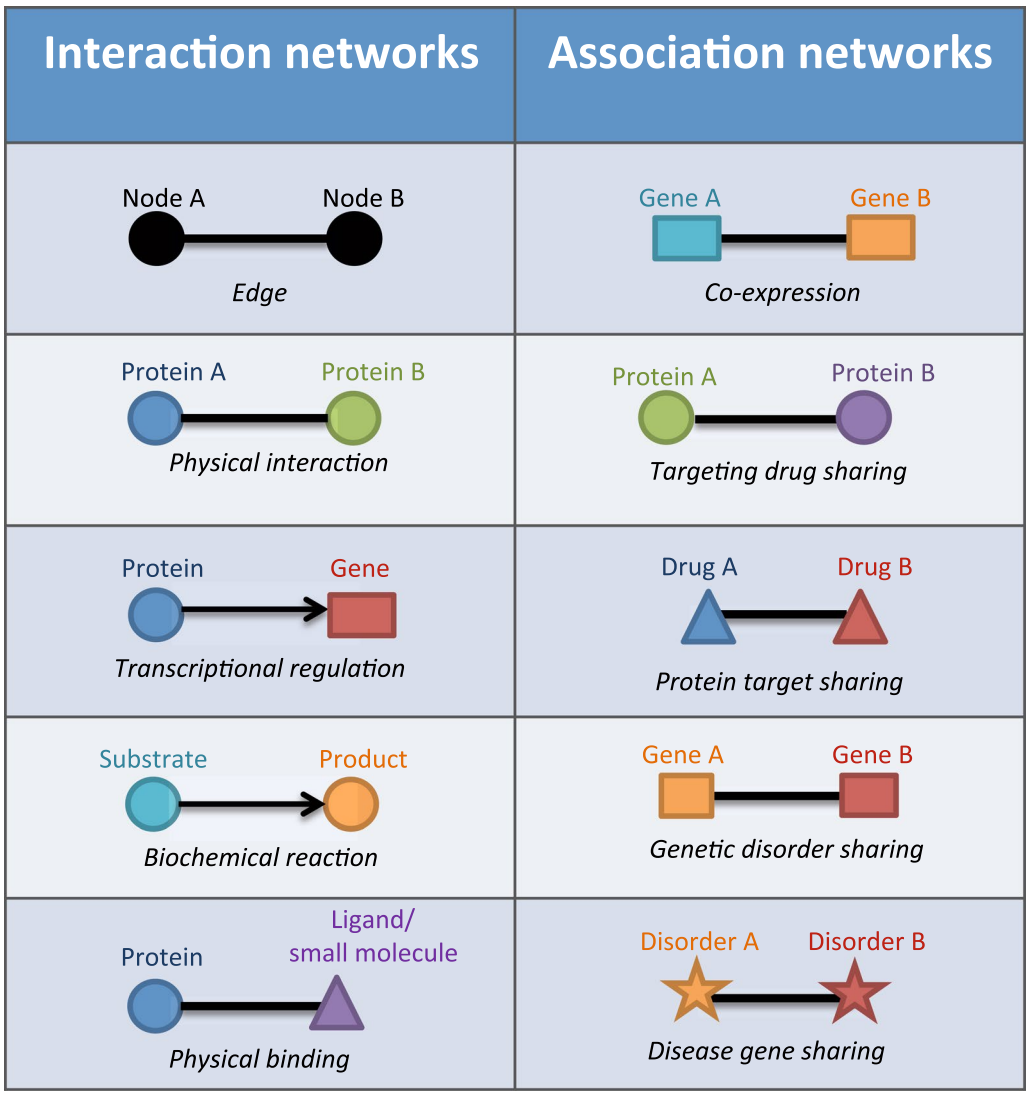


Fig. 1 Interaction networks (*Left*) represent direct interactions between biological molecules (e.g. transcripts, proteins, and ligands). The interactions represented include direct physical interaction (e.g. protein–protein, and gene regulatory networks) or transition (e.g. metabolic network). Association networks (*Right*) represent biological molecules that are linked based on their shared and/or common properties (e.g. co-expression)

Metabolic networks differs from other networks described earlier in the sense that the edges between two nodes (metabolites) do not represent physical contacts, but instead biochemical reactions that convert one metabolite to another. Recent studies have reconstructed and explored genome-scale metabolic networks in pathogenic microbes including *Staphylococcus aureus* [47], *M. tuberculosis* [48], as well as in human hosts [49]. These analyses may lead to a better understanding of host-pathogen interactions, and could aid in the design of drugs that specifically target the metabolic pathways of microbes and cause minimal interference with those of the hosts.

Association networks

Networks can also be used to visualize and summarize the overlap in expression profiles for thousands of transcripts/proteins obtained from high-throughput methods, such as expression microarray, RNA-seq, or short-gun proteomics [50]. In co-expression networks, two or more genes are linked if their products (mRNAs or proteins) exhibit similar expression profiles, with the strength/thickness of the edges proportional to how often the two transcripts are expressed at the same time and/or place [51, 52]. Co-expression networks are widely used as a starting point for inferring the cellular functions of uncharacterized genes, as in many cases, genes with

related functions show overlapping expression patterns [53]. New disease markers can be discovered from clusters of genes that are co-expressed with known disease-associated genes, as they frequently show differential expression between the normal and diseased populations [54–57].

Other association networks include drug target-protein networks [44], where each node is a protein and two proteins are linked if they are targeted by the same compounds. These networks can be computationally derived from the drug-target network described in the previous section. It provides a complementary protein-centric view by focusing on the proteins that are often co-targeted, and might be involved in related pathways. Conversely, two or more drugs can be linked in a network based on common properties, such as targeting specific proteins or side effects. It has been shown that documented adverse side effects could be used to infer molecular drug-target interactions [58]. This type of network has the potential to predict whether or not existing and routinely used drugs have additional unknown off-targets, allowing for these drugs to be candidates for additional, distinct therapeutic categories. Illustrations of the potential of alternative uses for current drugs are sildenafil, losartan, and fenofibrate. Sildenafil (e.g. Viagra®, Pfizer Incorporated) was initially developed to treat angina, but a side effect (prolong penile erection) discovered during clinical trial has become its main use. The antihypertensive drug losartan blocks angiotensin II type 1, and is now a candidate drug for preventing aortic aneurysm complications in Marfan syndrome patients, through reduction of TGF-β activitiy [59, 60]. Fenofibrate, a drug mainly used for controlling cholesterol levels in cardiovascular patients, has also been shown to suppress growth of hepatocellular carcinoma [61].

Global disease networks offer a useful insight into how human disorders are related. In the "human disease network" [62], disease nodes are connected if they share at least one gene with mutations associated with both diseases. Complementarily, the gene-centric version of this network comprises nodes of disease genes, linked if they are associated with the same disorders. Such networks not only represent a framework to visualize all known disease genotype-phenotype associations, but also reveal that human diseases are much more genetically related than previously appreciated [63]. This is highlighted by a gigantic network comprising over 500 interconnected human diseases [7].

What can we learn from networks and their properties?

In addition to being a framework for visualizing and documenting all the known relationships between nodes, earlier analyses of large-scale networks from high-throughput studies have revealed many interesting biologically relevant properties, which cannot be obtained by studying genes and proteins individually [64–66]. One of the most frequently observed properties of biological networks is the connectivity distribution that follows a power-law distribution, known as "scale-free networks". This pattern of connections, also known as the "small world property", has also been extensively studied for their statistical features in different types of networks, including social networks, scientific collaboration networks, and the World Wide Web [67–72]. In brief, a scale-free network consists of a small number of "hubs", i.e. nodes that are connected to a larger number of other nodes, through different types of interactions aforementioned. In contrast to hubs, the majority of nodes in the network have much fewer connections. Several studies have documented similar observation for biological networks, including protein–protein interaction networks [6, 17, 73] and metabolic networks [15, 74].

Because of their connectivity distribution, scale-free networks are robust against random deletion of nodes. That is, the connections between a node and most other nodes remain intact, if nodes are removed randomly. In contrast, scale-free networks quickly become non-functional if hubs are targeted. Earlier studies have shown that many pathogenic organisms have evolved to target the central components (i.e. hubs) of a human protein interaction network, and quickly disrupt various cellular functions, including the immune response [75, 76]. Similarly, one would expect drugs that specifically inhibit the central components of the regulatory circuits in a pathogen will rapidly disrupt their homeostatic processes, and thus efficiently eliminate them. As a result, these hubs from pathogenic organisms could be promising candidates for novel drugs. Network connectivity distribution is one of the better-studied areas, and a number of insightful reviews and analyses are available [77, 78].

Another interesting example of biological network properties are the network motifs, which are sets of well-defined interconnection patterns between nodes [19]. These connectivity patterns, or network sub-circuits, recur in biological networks at a frequency significantly higher than in randomized networks [79–81], signifying their important roles as building blocks for the large-scale organization of interactions. The patterns and proportions of sub-circuits used in different networks are distinct, depending on the functionality required under different conditions. Interestingly, it has been shown in a yeast transcription regulatory network that sub-network structures, facilitating fast signal propagation (e.g. single-inputs), are more frequently employed to respond to external stressors and sudden environmental changes (e.g. DNA damage or diauxic shift), because a rapid

response is required against the stressors. In contrast, motifs that buffer spurious inputs or only respond to persistent signals (e.g. feed-forward-loops) are more suitable for analysis of normal growth stages (e.g. sporulation) [18, 82].

Applications of network biology in translational medicine

Disease network and drug discovery

Using a transistor radio as an analog of a biological system, Yuri Lazebnik described how a biologist would fix a broken radio, assuming no prior knowledge of how the radio components were wired together [83]. A traditional biological approach would involve removing (gene knockout, mutagenesis) each part of a functioning radio and track the changes in performance (phenotype). However, the human "radios" are different and repeating this process on all the components would generate an enormous amount of data, some of which may be redundant or contradictory. In contrast, a typical engineering approach would involve systematic reconstruction of a component diagram from a normal radio (e.g. regulatory network), and compare the broken radios with the normal reference. Can a similar problem-solving mindset help expedite advances in biomedical research?

If regulatory circuits that control biological activities in a human body can be represented using a complex network, then a diseased state would be expected to occur when the normal state of the network is perturbed. Failure of key components (e.g. mutations in hub genes in genetic diseases) or external stimuli (e.g. invasion of pathogens in infectious diseases) would lead to loss of network integrity. Diseased perturbations can occur at different regulatory levels, as illustrated in Fig. 2. Firstly, the absence or malfunction in important network components can lead to diseases, such as the loss of a particular gene. The absence of *TBX1*, in 22q11.2 deletion syndrome (DiGeorge syndrome) is responsible for the majority of characteristic features of this disease [84] (Fig. 2a, the absence of node is illustrated in red). Similarly, inappropriate levels of gene expression can cause disorders (Fig. 2b, altered node size). For example, specific mutations in the *FGFR3* gene result in an overactive receptor and lead to the short stature phenotype observed in achondroplasia [85]. Some diseased states can be explained by mis-regulation of the interactions between key components of the network (Fig. 2c, missing edge), as well as mis-direction (Fig. 2d, mis-directed edge) or strength (Fig. 2e, altered edge's thickness) of interactions. The diseases that can be linked to erroneous interactions include neurodegenerative and neurodevelopmental diseases, genetic disorders, and cancers. In these cases, mutations in multiple relevant genes lead to abnormal protein interactions, and disrupt networks (see [29, 30, 36, 37] for details).

Some of the long-standing challenges in drug discovery are lack of specificity, high incidence of adverse effects, and unpredicted toxicities of new therapeutic compounds [86]. As a result, modern-day drug discovery employs more targeted approaches, such as virtual screening and structure-based drug design to complement conventional in vitro high-throughput screening [46, 87]. These new approaches rely on an accurate global understanding of the mechanisms of diseases. Comprehensive understanding of the network and regulatory circuit for a particular disease process would help to identify network hubs with the potential to be novel drug targets.

A network model of cancers

In the past decades, chemotherapy had been the backbone for systemic treatment of cancers. When administered to patients, these drugs target rapidly dividing cells but lack specificity. Survival of both cancer cells and normal, rapidly growing cells are impaired, resulting in side effects such as bone marrow suppression and hair loss, due to toxicity toward bone marrow cells and hair follicles, respectively. With recent advances in molecular biology and genetics, several genetic mutations and other alterations have been described for various cancers, and these changes specific to cancer cells have become an attractive target for novel therapies. The concept of "driver" and "passenger" mutations in carcinogenesis is comparable to hubs and peripheral nodes in a network, whereby a subset of somatic alterations present in each tumor is a driver of the oncogenic process [88]. Acting as a complex network hub, these driver mutations promote cancer cell survival, resistance to apoptosis, and lead to carcinogenesis (so-called "oncogene addiction"). This idea is supported by successful identification of new cancer fusion drivers from the network hubs and their partners, as the fusion mutation can lead to functional de-regulation of multiple genes and pathways [89]. Inhibition of the driver mutation has the potential to induce cell death, and thus becomes a strong candidate for targeted therapy [90]. As cancer cells are addicted to this driver mutation, specifically blocking these hubs would theoretically be more effective and less toxic compared to conventional chemotherapy.

To date, many targeted therapies have been approved as a standard of care in various cancers with additional clinical studies underway. Identification of a true driver; however, remains one of the biggest challenges. Pathogenesis of cancer development is usually complex and involves several molecules and pathways. Therefore, targeting one particular molecule or pathway might not be effective, as cancer cells may utilize alternative pathways

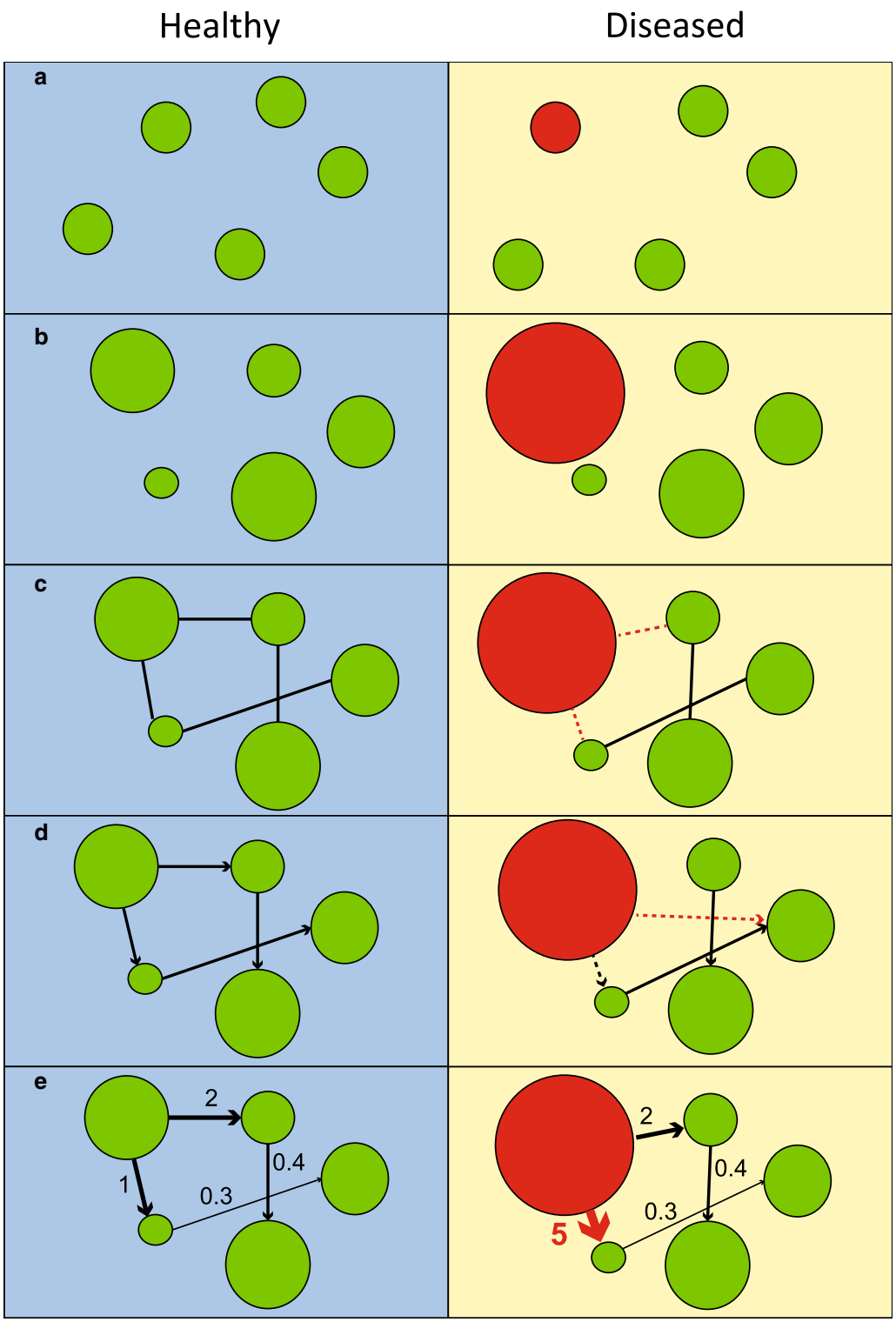


Fig. 2 Biological networks of healthy (*left panel*) and diseased (*right panel*) individuals. Biological components in healthy individuals are represented as *green nodes* in a network. Pathological perturbation, represented by *red nodes* that lead to morbidity, can occur at different stages of the regulation of key components: **a** presence and absence of key component (*green* for presence and *red* for absence), **b** mis-regulated gene expression, leading to over- or under-expression (*node sizes* represent expression levels), **c** absence or erroneous interactions with interacting partners (*dotted lines* represent erroneous interactions), **d** mis-regulated directions (*mis-directed arrows*), or **e** strengths of interactions (*thicknesses of arrows* and *accompanying numbers* denote interaction strengths)

to promote cell survival. Additionally, with the advent of next-generation sequencing, the previously well-accepted but unproven concept of tumor genetic heterogeneity has been solidly confirmed [91]. Sequential use of more than one targeted cancer therapy to finish off resistant clones, such as in the case of tumor recurrence, is likely to become a trend in cancer genomic medicine [92].

Breast cancer network: mechanisms of resistance

The regulatory network in breast cancer is a particularly interesting case study, due to its heterogeneous histological and molecular features, and clinical manifestations that lead to multiple molecular sub-types. Based on gene expression profiling, breast cancer can be categorized into four main molecular sub-types: (i) basal-like breast cancer (mainly estrogen-receptor (ER)-negative, progesterone-receptor (PR)-negative, and human epidermal growth factor receptor 2 (HER2)-negative); (ii) luminal-A cancer (ER-positive or ER+, and histologically low-grade); (iii) luminal-B cancer (ER+ and histologically high-grade); and (iv) HER2-positive (HER2+) cancer (over-expression and/or amplification of HER2). Each molecular sub-type has a distinct course of disease progression and responds differently to specific treatments, including endocrine therapy, anti-HER2 drugs and cytotoxic chemotherapy [93].

As shown in Fig. 3, ER and HER2 can be considered as hubs of the breast cancer network. The ER+ breast cancer cells depend on activation of ER by estrogen, a sex steroid hormone. ER acts as a transcription factor in the nucleus when bound by estrogen in the genomic (nuclear) pathway, resulting in tumor cell proliferation [94]. The signal can also be activated through the non-genomic (non-nuclear) pathway, where estrogen binds to membrane-associated ER. Endocrine therapy against the ER hubs is one of the cornerstones of treatment for ER+/HER2- breast cancers (luminal-A and B) [95]. The predominant endocrine therapies are a selective ER modulator (SERM), an aromatase inhibitor (AI), and selective ER down-regulators (SERD), such as tamoxifen, anastrozole, and fulvestrant [96].

HER2, a member of the epidermal growth factor receptor tyrosine kinase family, is a hub in the HER2+ breast cancer network. Over-expressed and/or amplified *HER2* is found in approximately 20–30% of invasive breast cancers [97]. HER2 activates intracellular signaling cascades, leading to tumor cell proliferation. Inhibition of HER2 through the use of anti-HER2 drugs significantly prolongs survival in HER2+ breast cancer patients. Currently, several anti-HER2 drugs are FDA-approved for HER2+ breast cancer, including trastuzumab, lapatinib, pertuzumab, and trastuzumab emtansine (T-DM1). Resistance to each of these specific treatments has been observed, as well as interactions between the ER and HER2 hubs (Fig. 3) [94, 98]. Since ER+/HER2+ tumor cells depend on both hubs, endocrine therapy alone cannot completely inhibit signals with tumor cell proliferation continuing to be activated through HER2 (so-called "cross-talk"). This has been identified as a primary mechanism of resistance in ER+/HER2+ breast cancer patients with a low response to endocrine therapy. With a better understanding of global gene regulation networks and the interplay between the two hubs, a combined treatment of endocrine therapy and anti-HER2 drugs was proposed. Several phase 3 clinical studies have already demonstrated increased efficacy of endocrine therapy in the ER+/HER2+ breast cancer when combined with anti-HER2 drugs [99–101].

On the other hand, ER+/HER− breast cancer does not depend on the HER2 hub, and is thus usually responsive to the first line endocrine therapy. However, resistance can still occur leading to less effective endocrine therapy. Blocking the ER hub with any endocrine therapy would inhibit only the genomic pathway, but not the non-genomic pathway where abnormal activation of the PI3K/Akt/mTOR pathway by somatic mutations can result in either *de novo* or acquired endocrine therapy resistance [102, 103]. Understanding this relationship has led to a second line of endocrine therapy using mTOR inhibitors. A large phase 3 clinical study of metastatic ER+/HER2− breast cancer patients, who failed the first line AI treatment, reported longer progression-free survival in a group treated with a combination of an mTOR inhibitor and another different AI [104, 105].

Having a comprehensive understanding of the interactions between network components of specific disease should lead to improved efficacy in treatments, similar to those elucidated using the breast cancer model above. Indeed, a number of groups have already begun utilizing network biology to address different aspects of cancers with the goal to improve diagnosis and treatment. A model to identify genes potentially associated with high risks of breast cancer has been developed by integrating data from co-expression, biochemical, and protein interaction networks. Using this model, Pujana and coworkers successfully identified Hyaluronan Mediated Motility Receptor (*HMMR*), a hub of the integrated network, as a novel high risk associated locus [31]. The gene regulatory network for breast cancer has also been constructed [106]. Taylor and colleagues merged spatial gene expression information with the protein interaction network to highlight the interactions that are active in specific tissues, where the interacting partners are also co-expressed [107]. This work also revealed the loss of key interactions between the network hubs, such as *BRCA1* and their binding partners, in patients who died of breast cancer due to mis-regulation

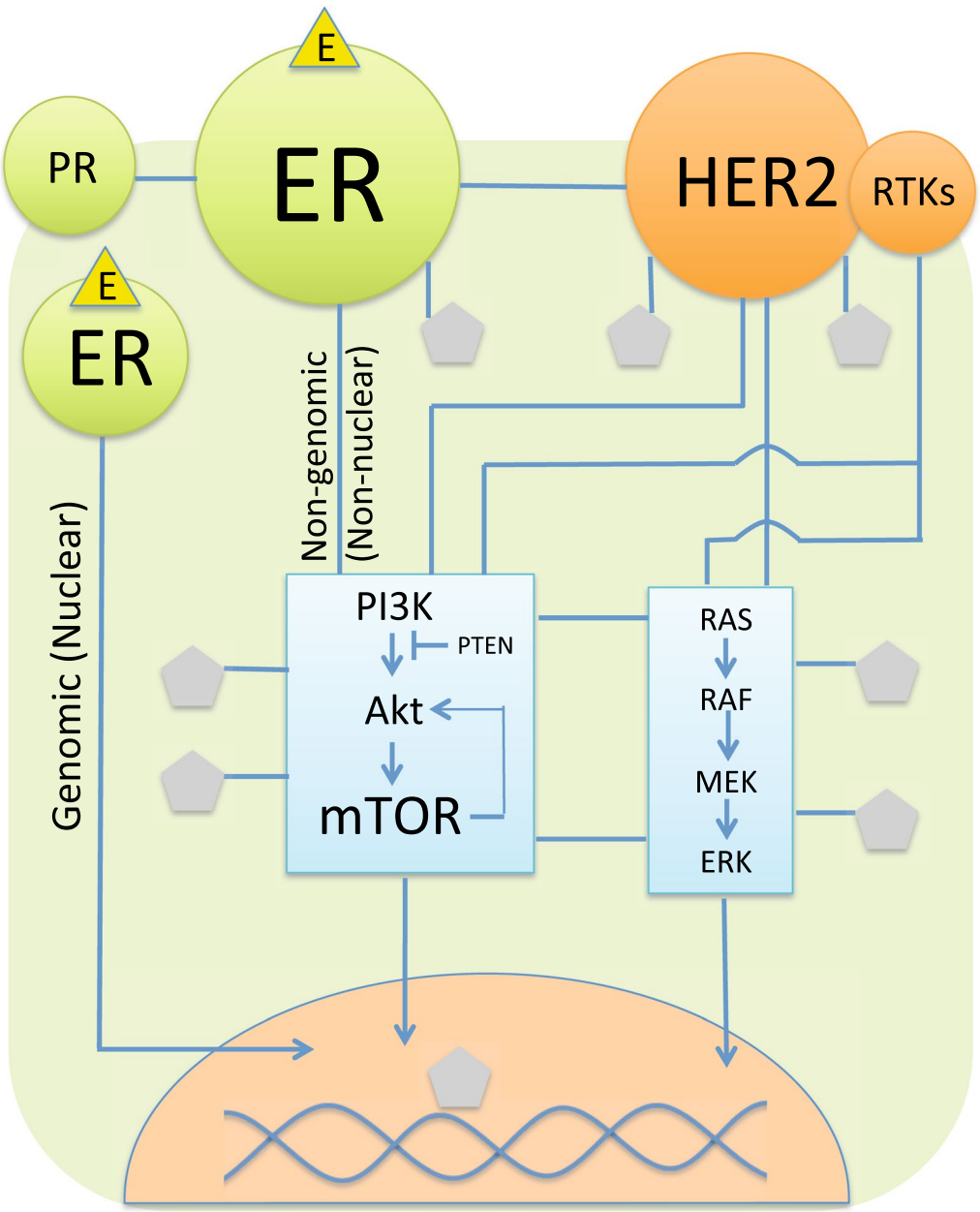


Fig. 3 A simplified diagram of the therapeutic breast cancer network. The main targetable hubs are ER and HER2 receptor. The PI3K/Akt/mTOR hub was relatively recently identified to be the common mechanism of targeted therapy resistance. *Circles* and *rectangles* represent cellular receptors and signaling pathways, respectively. The *pentagons* represent other unspecified molecules interacted with the hubs. *Arrows* represent the directions of signals. (*E* estrogen, *ER* estrogen receptor, *PR* progesterone receptor, *HER2* HER2 receptor, *RTKs* receptor tyrosine kinases)

of the partner proteins. In contrast, the expression of hubs and their partners were strongly correlated in surviving patients. The complexity of the disease network is not only restricted to the gene–gene and gene-drug interactions, but also hinges upon the interactions between disease/drug and the host (i.e. genetic background of the patients), as we discuss in the next section.

From individual network to personalized medicine

As we are approaching the so-called personalized and precision medicine era, where does network biology fit in the picture? Figure 4 depicts our view on how networks can be an important tool to help clinicians understand the physiological complexity of individual humans, predict possible failure of certain components that may lead

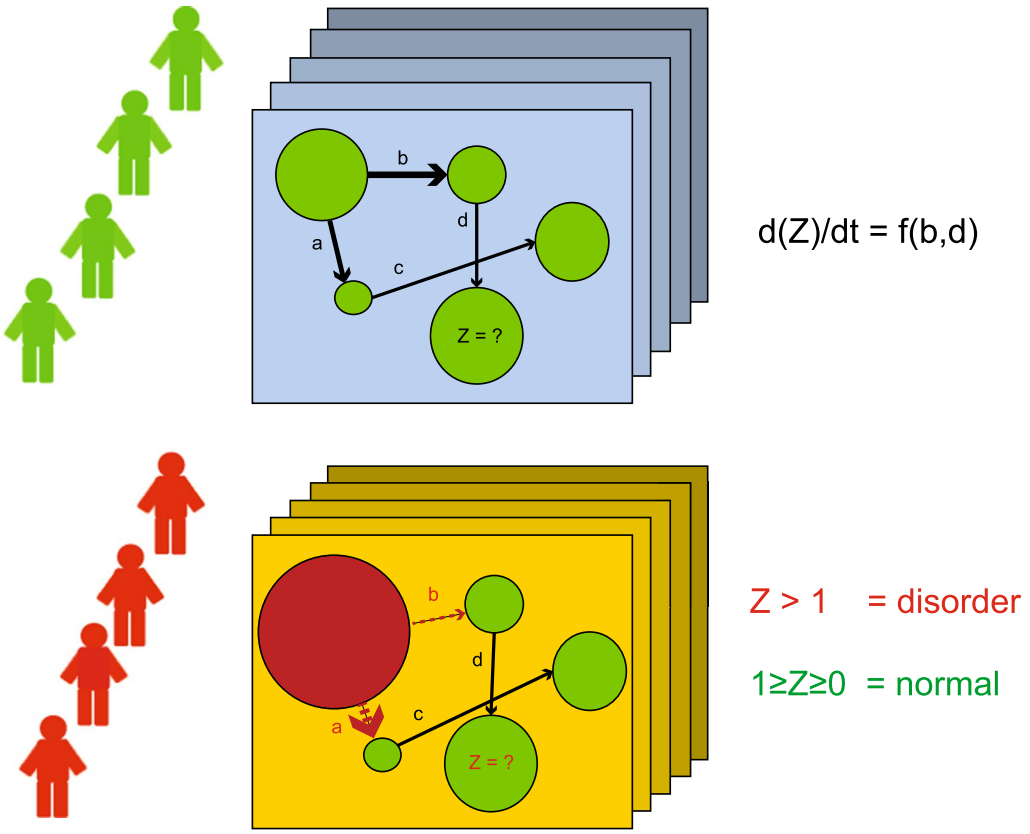


Fig. 4 Healthy (*top panel*) and diseased (*bottom panel*) individual networks. Healthy individuals might show slight variations in their individual networks, which also differ over time. However, diseased networks are expected to show greater disparity than that between healthy individuals. In the example shown, the network component *Z* is controlled by its upstream components through the interactions of *b* and *d* (the molecule Z is a function of *b* and *d*). If the expression of *Z* is greater than a defined limit (e.g. 1 in this case), morbidity can be predicted (d(Z)/dt: change of expression level of molecule Z over time)

to morbidity, and deduce the most suitable preventative and treatment plans for individual patients. Genetic variation between human individuals is estimated to be less than 1% of the human genome, but through sophisticated regulation of genes and other genetic elements, this small amount of genetic variation accounts for much greater differences in terms of our appearance, intellect, and health [108]. On top of genomes, which encode individual sets of gene products (e.g. proteins, mRNA), individual networks represent the unique interplay between different components in each patient. Understanding the extent of variations between individual networks may allow clinicians to statistically and quantitatively distinguish normal variations in healthy individuals (Fig. 4, upper panel) from critical perturbations that lead to diseases and disorders (Fig. 4, lower panel). Network biology enables researchers to assess multiple components that do not show distinguishable differences between healthy individuals and those with cancers, but are collectively dysfunctional in cancers. A sub-network in which overall activity can be discriminated between patients versus controls has been shown to be a more reproducible prognostic marker of diseases than individual genes in the sub-network, which are not significantly differentially expressed [109, 110].

Single nucleotide polymorphisms (SNPs) and other genetic variations add another dimension of disease-host interaction to disease networks. SNPs can provide clinicians with a good indication on how likely an individual might be to develop certain genetic diseases, assuming that all genetic elements associated with diseases are eventually identified. In addition, networks of individuals can, in part, aid pharmacogenomic progress by explaining why the efficacy and toxicity profiles for the same drug may differ in each patient. For instance, tamoxifen is metabolized by *CYP2D6* and variations in this gene among individuals may affect the response to the drug [111].

No matter how comprehensive, a genetic map cannot capture environmental factors (e.g. lifestyle, contact with

pathogens) that heavily influence biochemical stages. Thus, outcomes for the interplay between genetics and environment may be absent in the analysis. Having a network that combines both the genetic variations and measurable biochemical outcomes, such as gene expression, should assist in turning conceptual ideas into more quantitative models, which in turn would enhance the accuracy of prognosis and predictions of disease progression in each patient (as demonstrated in Fig. 4). Such a complete individual network may not be possible in the near future; however, we start to see that the integration of genetic variations and biochemical outcomes (gene expression and protein interaction profiles) has utility in helping identify new disease-associated marker genes [110, 112, 113].

Thanks to considerable effort and resources the community has put into developing computational tools for biological network analysis, we are now well-equipped with a range of user-friendly software that can be employed to handle, visualize, and analyze large-scale datasets. Importantly, the tools that will be particularly useful for translational medical research need to be able to combine multiple layer datasets (e.g. genomics, transcriptomics, proteomics, and metabolomics) and/or heterogeneous datasets (e.g. from different platforms or formats) [3]. The most commonly known network analysis tools currently available are Cytoscape [114], NAViGaTOR [115], VisANT [116], CellDesigner [117], and the commercial software Ingenuity IPA (Ingenuity Systems Inc., Redwood City, CA). More recently introduced tools include NaviCell, which has been developed for online network visualization and curation [118], and BNOmics [119], which can be used for inference and visualization of Bayesian networks of large heterogeneous data. Comprehensive guides to network biology tools, as well as detailed discussion on their key features and functionality can be found in earlier review articles [3, 120].

Conclusions

Network biology provides an opportunity to image a clear global picture of drug-disease-host interactions and the biological complexity of diseases more easily from an unprecedented top-down vantage. This will allow a better understanding of the relationships between multiple genes and other biological entities, as well as identify the missing links in our knowledge. These strategies are required to fully grasp the intricacies of diseases, which cannot be obtained by studying an individual or a smaller set of genes. The complexity of the therapeutic networks is ever-growing, and many new nodes are being discovered every day. In the future, some of these nodes may become new hubs for targeted therapy.

Abbreviations

ER: estrogen receptor; FDA: Food and Drug Administration; GRN: gene regulatory network; HER2: human epidermal growth factor receptor 2; HMMR: hyaluronan mediated motility receptor; NGS: next-generation sequencing; PR: progesterone receptor; SNP: single nucleotide polymorphisms.

Authors' contributions

NJ, NA and VC conceived the concept of the review and figures. NJ, SB, NN and VC wrote the manuscript. MC, NA and JS contributed the ideas and literature search. SB, NA and VC prepared the figures. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Author details

[1] Integrative Computational BioScience (ICBS) Center, Mahidol University, Nakhon Pathom, Thailand. [2] Program in Translational Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. [3] Department of Physiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand. [4] Division of Gastroenterology and Hepatology, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. [5] Medical Oncology Unit, Department of Medicine Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. [6] Department of Physiology, Faculty of Science, Mahidol University, Bangkok, Thailand. [7] Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand. [8] Laboratory of Biochemistry, Chulabhorn Research Institute, Bangkok, Thailand. [9] Systems Biology of Diseases Research Unit, Faculty of Science, Mahidol University, Bangkok, Thailand.

Acknowledgements

The authors thank Assoc. Prof. Laran Jensen, Department of Biochemistry, Mahidol University, for very helpful comments on the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

NJ is a recipient of TRF Research Scholar Fund (RSA5780065), Mahidol University-National Research Council of Thailand (Government Fiscal Year Budget) Fund, and the Research grants from the Ramathibodi Cancer Center. SB is a recipient of research assistant scholarship from Faculty of Medicine Ramathibodi Hospital and Faculty of Graduate Studies, Mahidol University. NN acknowledges the Talent Management Program, Mahidol University. VC acknowledges the TRF Grant for New Researcher (TRG5880067), and Faculty of Science, Mahidol University. The NJ and VC laboratories are supported by the Crown Property Bureau Foundation through Integrative Computational BioScience (ICBS) Center, Mahidol University.

Received: 23 August 2016 Accepted: 8 November 2016
Published online: 22 November 2016

References

1. Attur MG, Dave MN, Tsunoyama K, Akamatsu M, Kobori M, Miki J, Abramson SB, Katoh M, Amin AR. "A system biology" approach to bioinformatics and functional genomics in complex human diseases: arthritis. Curr Issues Mol Biol. 2002;4(4):129–46.
2. Ideker T, Lauffenburger D. Building with a scaffold: emerging strategies for high-to low-level cellular modeling. Trends Biotechnol. 2003;21(6):255–62.
3. Chuang HY, Hofree M, Ideker T. A decade of systems biology. Annu Rev Cell Dev Biol. 2010;26:721–44.
4. Kitano H. Systems biology: a brief overview. Science. 2002;295(5560):1662–4.
5. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, et al. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006;440(7084):631–6.

6. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al. High-quality binary protein interaction map of the yeast interactome network. Science. 2008;322(5898):104–10.
7. Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. Cell. 2011;144(6):986–98.
8. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, et al. A proteome-scale map of the human interactome network. Cell. 2014;159(5):1212–26.
9. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, et al. Transcriptional regulatory code of a eukaryotic genome. Nature. 2004;431(7004):99–104.
10. Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008;132(6):1049–61.
11. Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier LW, Janette J, Jiang L, et al. Comparative analysis of regulatory information and circuits across distant species. Nature. 2014;512(7515):453–6.
12. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science. 2005;307(5715):1621–5.
13. Bhalla US, Ram PT, Iyengar R. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science. 2002;297(5583):1018–23.
14. Li L, Tibiche C, Fu C, Kaneko T, Moran MF, Schiller MR, Li SS, Wang E. The human phosphotyrosine signaling network: evolution and hotspots of hijacking in cancer. Genome Res. 2012;22(7):1222–30.
15. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature. 2000;407(6804):651–4.
16. Oberhardt MA, Goldberg JB, Hogardt M, Papin JA. Metabolic network analysis of *Pseudomonas aeruginosa* during chronic cystic fibrosis lung infection. J Bacteriol. 2010;192(20):5534–48.
17. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004;430(6995):88–93.
18. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature. 2004;431(7006):308–12.
19. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet. 2002;31(1):64–8.
20. Kohestani H, Giuliani A. Organization principles of biological networks: an explorative study. Biosystems. 2016;141:31–9.
21. Emmert-Streib F, Glazko GV. Network biology: a direct approach to study biological function. Wiley Interdiscip Rev Syst Biol Med. 2011;3(4):379–91.
22. Zhang H, Gustafsson M, Nestor C, Chung KF, Benson M. Targeted omics and systems medicine: personalising care. Lancet Respir Med. 2014;2(10):785–7.
23. Hood L. Systems biology and p4 medicine: past, present, and future. Rambam Maimonides Med J. 2013;4(2):e0012.
24. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56–68.
25. Clermont G, Auffray C, Moreau Y, Rocke DM, Dalevi D, Dubhashi D, Marshall DR, Raasch P, Dehne F, Provero P. Bridging the gap between systems biology and medicine. Genome Med. 2009;1(9):88.
26. Michor F, Liphardt J, Ferrari M, Widom J. What does physics have to do with cancer? Nat Rev Cancer. 2011;11(9):657–70.
27. Waaijers S, Koorman T, Kerver J, Boxem M. Identification of human protein interaction domains using an ORFeome-based yeast two-hybrid fragment library. J Proteome Res. 2013;12(7):3181–92.
28. Jirawatnotai S, Hu Y, Michowski W, Elias JE, Becks L, Bienvenu F, Zagozdzon A, Goswami T, Wang YE, Clark AB, et al. A function for cyclin D1 in DNA repair uncovered by protein interactome analyses in human cancers. Nature. 2011;474(7350):230–4.
29. Lim J, Hao T, Shaw C, Patel AJ, Szabo G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell. 2006;125(4):801–14.
30. Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA, et al. Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. Nat Commun. 2014;5:3650.
31. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat Genet. 2007;39(11):1338–49.
32. Nibbe RK, Koyuturk M, Chance MR. An integrative -omics approach to identify functional sub-networks in human colorectal cancer. PLoS Comput Biol. 2010;6(1):e1000639.
33. Hajingabo LJ, Daakour S, Martin M, Grausenburger R, Panzer-Grumayer R, Dequiedt F, Simonis N, Twizere JC. Predicting interactome network perturbations in human cancer: application to gene fusions in acute lymphoblastic leukemia. Mol Biol Cell. 2014;25(24):3973–85.
34. Charoensawan V, Adryan B, Martin S, Sollner C, Thisse B, Thisse C, Wright GJ, Teichmann SA. The impact of gene expression regulation on evolution of extracellular signaling pathways. Mol Cell Proteomics. 2010;9(12):2666–77.
35. Keith BP, Robertson DL, Hentges KE. Locus heterogeneity disease genes encode proteins with high interconnectivity in the human protein interaction network. Front Genet. 2014;5:434.
36. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, Peng J, Weile J, Karras GI, Wang Y, et al. Widespread macromolecular interaction perturbations in human genetic disorders. Cell. 2015;161(3):647–60.
37. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi AL. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015;347(6224):1257601.
38. Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. BioEssays. 1998;20(5):433–40.
39. Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, et al. A gene-centered *C. elegans* protein-DNA interaction network. Cell. 2006;125(6):1193–205.
40. Liu X, Huang J, Chen T, Wang Y, Xin S, Li J, Pei G, Kang J. Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. Cell Res. 2008;18(12):1177–89.
41. Huang TS, Li L, MoalimNour L, Jia D, Bai J, Yao Z, Bennett SA, Figeys D, Wang L. A regulatory network involving beta-catenin, E-cadherin, PI3K/Akt, and slug balances self-renewal and differentiation of human pluripotent stem cells in response to Wnt signaling. Stem Cells. 2015;33(5):1419–33.
42. Scharer CD, McCabe CD, Ali-Seyed M, Berger MF, Bulyk ML, Moreno CS. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. Cancer Res. 2009;69(2):709–17.
43. Wang Q, Li W, Liu XS, Carroll JS, Jänne OA, Keeton EK, Chinnaiyan AM, Pienta KJ, Brown M. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol Cell. 2007;27(3):380–92.
44. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. Nat Biotechnol. 2007;25(10):1119–26.
45. Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu Rev Pharmacol Toxicol. 2012;52:505–21.
46. Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, Huang XP, Norval S, Sassano MF, Shin AI, Webster LA, et al. Automated design of ligands to polypharmacological profiles. Nature. 2012;492(7428):215–20.
47. Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BO. Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proc Natl Acad Sci USA. 2016;113(26):E3801–9.
48. Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. Mol Syst Biol. 2010;6:422.
49. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci USA. 2007;104(6):1777–82.

50. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol. 2007;25(1):117–24.
51. Ruan J, Dean AK, Zhang W. A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC Syst Biol. 2010;4:8.
52. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science. 2003;302(5643):249–55.
53. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, et al. Functional discovery via a compendium of expression profiles. Cell. 2000;102(1):109–26.
54. Xu Y, Duanmu H, Chang Z, Zhang S, Li Z, Liu Y, Li K, Qiu F, Li X. The application of gene co-expression network reconstruction based on CNVs and gene expression microarray data in breast cancer. Mol Biol Rep. 2012;39(2):1627–37.
55. Zhang J, Xiang Y, Ding L, Keen-Circle K, Borlawsky TB, Ozer HG, Jin R, Payne P, Huang K. Using gene co-expression network analysis to predict biomarkers for chronic lymphocytic leukemia. BMC Bioinformatics. 2010;11(Suppl 9):S5.
56. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature. 2010;463(7279):318–25.
57. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119–24.
58. Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, Lavan P, Weber E, Doak AK, Cote S, et al. Large-scale prediction and testing of drug activity on side-effect targets. Nature. 2012;486(7403):361–7.
59. Cohn RD, van Erp C, Habashi JP, Soleimani AA, Klein EC, Lisi MT, Gamradt M, ap Rhys CM, Holm TM, Loeys BL. Angiotensin II type 1 receptor blockade attenuates TGF-β–induced failure of muscle regeneration in multiple myopathic states. Nat Med. 2007;13(2):204–10.
60. Williams A, Davies S, Stuart A, Wilson D, Fraser A. Medical treatment of Marfan syndrome: a time for change. Heart. 2008;94(4):414–21.
61. Yamasaki D, Kawabe N, Nakamura H, Tachibana K, Ishimoto K, Tanaka T, Aburatani H, Sakai J, Hamakubo T, Kodama T. Fenofibrate suppresses growth of the human hepatocellular carcinoma cell via PPARα-independent mechanisms. Eur J Cell Biol. 2011;90(8):657–64.
62. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci USA. 2007;104(21):8685–90.
63. Hidalgo CA, Blumm N, Barabasi AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol. 2009;5(4):e1000353.
64. Cho DY, Kim YA, Przytycka TM. Network biology approach to complex diseases. PLoS Comput Biol. 2012;8(12):e1002820.
65. Furlong LI. Human diseases through the lens of network biology. Trends Genet. 2013;29(3):150–9.
66. Marcotte E, Boone C, Babu MM, Gavin A-C. Network biology editorial 2013. Mol BioSyst. 2013;9(7):1557–8.
67. Newman ME. The structure of scientific collaboration networks. Proc Natl Acad Sci USA. 2001;98(2):404–9.
68. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999;286(5439):509–12.
69. Strogatz SH. Exploring complex networks. Nature. 2001;410(6825):268–76.
70. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998;393(6684):440–2.
71. Ipsen M, Mikhailov AS. Evolutionary reconstruction of networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2002;66(4 Pt 2):046109.
72. Milgram S. The small world problem. Psychol Today. 1967;2:60.
73. Martin S, Sollner C, Charoensawan V, Adryan B, Thisse B, Thisse C, Teichmann S, Wright GJ. Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. Mol Cell Proteomics. 2010;9(12):2654–65.
74. Wagner A, Fell DA. The small world inside large metabolic networks. Proc Biol Sci. 2001;268(1478):1803–10.
75. Brodsky IE, Medzhitov R. Targeting of immune signalling networks by bacterial pathogens. Nat Cell Biol. 2009;11(5):521–6.
76. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali TM, Sobral BW. The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. PLoS ONE. 2010;5(8):e12089.
77. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5(2):101–13.
78. Stumpf MP, Porter MA. Mathematics. Critical truths about power laws. Science. 2012;335(6069):665–6.
79. Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007;8(6):450–61.
80. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010;468(7326):911–20.
81. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002;298(5594):824–7.
82. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol. 2004;14(3):283–91.
83. Lazebnik Y. Can a biologist fix a radio?-or, what I learned while studying apoptosis. Cancer Cell. 2002;2(3):179–82.
84. Gao S, Moreno M, Eliason S, Cao H, Li X, Yu W, Bidlack FB, Margolis HC, Baldini A, Amendt BA. TBX1 protein interactions and microRNA-96-5p regulation controls cell proliferation during craniofacial and dental development: implications for 22q11.2 deletion syndrome. Hum Mol Genet. 2015;24(8):2330–48.
85. Di Rocco F, Biosse Duplan M, Heuze Y, Kaci N, Komla-Ebri D, Munnich A, Mugniery E, Benoist-Lasselin C, Legeai-Mallet L. FGFR3 mutation causes abnormal membranous ossification in achondroplasia. Hum Mol Genet. 2014;23(11):2914–25.
86. Bantscheff M, Scholten A, Heck AJ. Revealing promiscuous drug–target interactions by chemical proteomics. Drug Discov Today. 2009;14(21):1021–9.
87. Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov. 2010;9(4):273–6.
88. Torkamani A, Verkhivker G, Schork NJ. Cancer driver mutations in protein kinase genes. Cancer Lett. 2009;281(2):117–27.
89. Wu CC, Kannan K, Lin S, Yen L, Milosavljevic A. Identification of cancer fusion drivers using network fusion centrality. Bioinformatics. 2013;29(9):1174–81.
90. Ma WW, Adjei AA. Novel agents on the horizon for cancer therapy. CA Cancer J Clin. 2009;59(2):111–37.
91. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
92. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. Br J Cancer. 2013;108(3):479–85.
93. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thurlimann B, Senn HJ. Tailoring therapies–improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. 2015;26(8):1533–46.
94. Osborne CK, Schiff R. Mechanisms of endocrine resistance in breast cancer. Annu Rev Med. 2011;62:233–47.
95. Goldhirsch A, Winer E, Coates A, Gelber R, Piccart-Gebhart M, Thürlimann B, Senn H-J, Albain KS, André F, Bergh J. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24(9):2206–23.
96. Wood AJ, Riggs BL, Hartmann LC. Selective estrogen-receptor modulators—mechanisms of action and application to clinical practice. N Engl J Med. 2003;348(7):618–29.
97. Hudis CA. Trastuzumab–mechanism of action and use in clinical practice. N Engl J Med. 2007;357(1):39–51.
98. Arpino G, De Angelis C, Giuliano M, Giordano A, Falato C, De Laurentiis M, De Placido S. Molecular mechanism and clinical implications of endocrine therapy resistance in breast cancer. Oncology. 2010;77(Suppl 1):23–37.

99. Argiris A, Wang CX, Whalen SG, DiGiovanna MP. Synergistic interactions between tamoxifen and trastuzumab (Herceptin). Clin Cancer Res. 2004;10(4):1409–20.
100. Johnston S, Pippen J, Pivot X, Lichinitser M, Sadeghi S, Dieras V, Gomez HL, Romieu G, Manikhas A, Kennedy MJ. Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor–positive metastatic breast cancer. J Clin Oncol. 2009;27(33):5538–46.
101. Kaufman B, Mackey JR, Clemens MR, Bapsy PP, Vaid A, Wardley A, Tjulandin S, Jahn M, Lehle M, Feyereislova A. Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomized phase III TAnDEM study. J Clin Oncol. 2009;27(33):5529–37.
102. Miller TW, Balko JM, Arteaga CL. Phosphatidylinositol 3-kinase and antiestrogen resistance in breast cancer. J Clin Oncol. 2011;29(33):4452–61.
103. Sabnis G, Goloubeva O, Jelovac D, Schayowitz A, Brodie A. Inhibition of the phosphatidylinositol 3-kinase/Akt pathway improves response of long-term estrogen-deprived breast cancer xenografts to antiestrogens. Clin Cancer Res. 2007;13(9):2751–7.
104. Wolff AC, Lazar AA, Bondarenko I, Garin AM, Brincat S, Chow L, Sun Y, Neskovic-Konstantinovic Z, Guimaraes RC, Fumoleau P. Randomized phase III placebo-controlled trial of letrozole plus oral temsirolimus as first-line endocrine therapy in postmenopausal women with locally advanced or metastatic breast cancer. J Clin Oncol. 2013;31(2):195–202.
105. Dees EC, Carey LA. Improving endocrine therapy for breast cancer: it's not that simple. J Clin Oncol. 2013;31(2):171–3.
106. Baselga J, Campone M, Piccart M, Burris HA III, Rugo HS, Sahmoud T, Noguchi S, Gnant M, Pritchard KI, Lebrun F. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. N Engl J Med. 2012;366(6):520–9.
107. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol. 2009;27(2):199–204.
108. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA. The sequence of the human genome. Science. 2001;291(5507):1304–51.
109. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007;3(1):140.
110. Leiserson MD, Vandin F, Wu H, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet. 2015;47(2):106–14.
111. Westbrook K, Stearns V. Pharmacogenomics of breast cancer therapy: an update. Pharmacol Ther. 2013;139(1):1–11.
112. Schubert M, Iorio F. Exploiting combinatorial patterns in cancer genomic data for personalized therapy and new target discovery. Pharmacogenomics. 2014;15(16):1943–6.
113. Barrenas F, Chavali S, Alves AC, Coin L, Jarvelin M-R, Jornsten R, Langston MA, Ramasamy A, Rogers G, Wang H. Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. Genome Biol. 2012;13(6):R46.
114. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
115. Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W, Devani B, Toch IL, Jurisica I. NAViGaTOR: network analysis, visualization and graphing toronto. Bioinformatics. 2009;25(24):3327–9.
116. Hu Z, Snitkin ES, DeLisi C. VisANT: an integrative framework for networks in systems biology. Brief Bioinform. 2008;9(4):317–25.
117. Matsuoka Y, Funahashi A, Ghosh S, Kitano H. Modeling and simulation using cell designer. Methods Mol Biol. 2014;1164:121–45.
118. Bonnet E, Viara E, Kuperstein I, Calzone L, Cohen DP, Barillot E, Zinovyev A. NaviCell Web service for network-based data visualization. Nucleic Acids Res. 2015;43(W1):W560–5.
119. Gogoshin G, Boerwinkle E, Rodin AS. New algorithm and software (BNOmics) for inferring and visualizing bayesian networks from heterogeneous big biological and genetic data. J Comput Biol. 2016;23:1–17. doi:10.1089/cmb.2016.0100.
120. Thomas S, Bonchev D. A survey of current software for network analysis in molecular biology. Hum Genomics. 2010;4(5):353–60.