Gene expression signature: a powerful approach for drug discovery in diabetes

Smithamol Sithara¹, Tamsyn M Crowley², Ken Walder¹ and Kathryn Aston-Mourney¹

¹Metabolic Research Unit, School of Medicine, Deakin University, Geelong, Australia
²School of Medicine, MMR, Bioinformatics Core Research Facility, Deakin University, Geelong, Australia

Abstract

Type 2 diabetes (T2D) is increasing in prevalence at an alarming rate around the world. Much effort has gone into the discovery and design of antidiabetic drugs; however, those already available are unable to combat the underlying causes of the disease and instead only moderate the symptoms. The reason for this is that T2D is a complex disease, and attempts to target one biological pathway are insufficient to combat the full extent of the disease. Additionally, the underlying pathophysiology of this disease is yet to be fully elucidated making it difficult to design drugs that target the mechanisms involved. Therefore, the approach of designing new drugs aimed at a specific molecular target is not optimal and a more expansive, unbiased approach is required. In this review, we will look at the current state of diabetes treatments and how these target the disease symptoms but are unable to combat the underlying causes. We will also review how the technique of gene expression signatures (GESs) has been used successfully for other complex diseases and how this may be applied as a powerful tool for the discovery of new drugs for T2D.

Introduction

Type 2 diabetes (T2D) is an increasing problem currently affecting the health and lifestyle of more than 422 million people worldwide (NCD-RisC 2016), a number that is expected to rise to 592 million by 2035 (Guariguata et al. 2014). T2D is characterised by two main defects, insulin resistance and beta-cell dysfunction that together contribute to the diabetic milieu.

Under normal conditions, when there is a rise in blood glucose, glucose is transported into the pancreatic beta-cell via glucose transporters where it undergoes glycolysis and signals for insulin to be secreted (Ashcroft & Rorsman 2012, Esguerra et al. 2014). However, under insulin-resistant conditions, typically associated with obesity, the increased levels of plasma glucose stimulate the beta-cells to produce more insulin by increasing both their secretory activity and mass (Butler et al. 2003, Weir & Bonner-Weir 2004). Therefore, even in the presence of insulin resistance, normoglycaemia can be maintained (Steil et al. 2001, Jetton et al. 2005, Nolan et al. 2006). However, when the beta-cell is also predisposed to failure, due to complex genetic and environmental factors, this continuous demand to increase the amount of insulin production can result in beta-cell failure, where beta-cell function and mass deteriorate resulting in hyperglycaemia (Leahy 2005). Importantly, the conversion of beta-cell adaptation to beta-cell deterioration can be a slow process...
as evidenced by the slow progression (usually over several years) of subjects with impaired glucose tolerance (IGT), a pre-diabetic state, to overt diabetes (Motala et al. 1993, Tuomilehto et al. 2001).

Multiple genetic and environmental factors are associated with T2D and are thought to contribute to the imbalance in tightly regulated metabolic processes, and this eventually leads to the complexity surrounding the pathogenesis of T2D. The genetic component is evident in family and population based studies in patients who are at risk of T2D (Lehtovirta et al. 2000, Grarup et al. 2007, Voight et al. 2010). Additionally, a number of large genome-wide association studies have identified risk loci that could be involved in the pathogenesis of T2D (Sandhu et al. 2007, Scott et al. 2007, Sladek et al. 2007, Zeggini et al. 2008, Saxena et al. 2012, DIAGRAM et al. 2014). However, although more than 80 loci have been identified, each of this is a common variant that only has a small effect to increase the risk of T2D (Andersen et al. 2016). Further, with most of these loci, the causal variant has not been identified (Andersen et al. 2016). Therefore, although we are slowly increasing our understanding of the genetic basis of T2D, we are still a long way away from having a complete view. In addition to the genetic factors, many environmental factors play important roles in the development of T2D. These factors include increased caloric and the nutrient composition of food intake, reduced energy expenditure, the in utero environment, gut microbiome, alterations in diurnal patterns and exposure to different chemicals (reviewed in Kahn et al. 2014). Not only are the genetic and environmental factors involved in this disease complex but also the cellular mechanisms are just as complex. Glucotoxicity, lipotoxicity, oxidative stress, endoplasmic reticulum stress, islet amyloidogenesis, inflammation, altered neuronal signalling and epigenetic reprogramming have all been shown to play a role in the pathophysiology of T2D (reviewed in Hull et al. 2004, McGee & Hargreaves 2010, Bensellam et al. 2012, Biden et al. 2014, Kahn et al. 2014). Therefore, it is clear that the causative and contributory genetic and environmental factors involved in T2D as well as the cellular mechanisms mediating their effects are all very complex, and we are still a long way from fully understanding this multifactorial interplay.

Given the complex nature of T2D, it stands to reason that any effective treatment will have to target multiple aspects of the underlying mechanisms causing the disease, which will be very difficult. One of the best ways to combat T2D is lifestyle interventions aimed at prevention. Lifestyle interventions including diet and increased exercise resulting in weight loss can successfully reduce the development of T2D in high-risk populations (Tuomilehto et al. 2001, Knowler et al. 2002). Further, lifestyle interventions after the development of T2D are also recommended as continual treatment strategies to improve glycaemic control and reduce complications (American Diabetes Association 2016, Royal Australian College of General Practitioners 2016). However, lifestyle changes alone cannot combat the epidemic of diabetes and despite these recommendations, and available therapies, not enough patients with type 2 diabetes mellitus achieve adequate glycaemic control (Mitka 2013). Below, we review the antidiabetic medications currently available and their possible limitations, highlighting the need for an unbiased drug discovery tool that can overcome the complex nature of T2D.

**Current therapeutic approaches for T2D**

The increasing prevalence of T2D has led to different approaches in the discovery of new therapeutic targets for the treatment of hyperglycaemia. Currently, there are numerous oral and injectable drugs available for the treatment of T2D. Although the main focus of these current treatments is to achieve glycaemic control and to delay the complications associated with the disease, achieving and maintaining long-term glycaemic control is challenging (Aston-Mourney et al. 2008). In addition, although they act to lower blood glucose, which is beneficial, they generally do not directly contribute to improving the underlying causes of T2D.

There are many classes of antidiabetic drugs available for the treatment of T2D including biguanides, thiazolidinediones, sulphonylureas, glucagon-like peptide-1 (GLP-1) modulators (GLP-1 mimetics and gliptins), α-glucosidase inhibitors, sodium-linked glucose transporter 2 (SGLT2) inhibitors and exogenous insulin, which are each described in Table 1. These drugs are given alone or in combination depending on the level of glycaemic control obtained by each individual.

**The need for new T2D therapies**

There is an ongoing search for the development of therapeutic agents with new mechanisms of action and
Table 1  Current antidiabetic drugs classes, action and limitations.

<table>
<thead>
<tr>
<th>Drug class</th>
<th>Mode of action</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biguanides</td>
<td>• Reduce hepatic gluconeogenesis (Inzucchi et al. 1998, Zhou et al. 2001), oxidation of fatty acids (Geerling et al. 2014) and glycogenolysis (Chu et al. 2000)</td>
<td>• Little effect on disease progression</td>
</tr>
<tr>
<td>Thiazolidinediones</td>
<td>• Increase glucose disposal via activating PPAR-γ to improve insulin signalling in liver, adipose tissue and skeletal muscles (Mags et al. 1998, Miyazaki et al. 2001, 2003, Kim et al. 2002)</td>
<td>• Antidiabetic effect declines over time (Weeks &amp; Lathrop 1995)</td>
</tr>
<tr>
<td>Sulphonylureas</td>
<td>• Directly increase insulin secretion via closing potassium-ATP channel (SUR-1 subunit) on beta-cells (Proks et al. 2002)</td>
<td>• Gastrointestinal side effects (Florez et al. 2010)</td>
</tr>
<tr>
<td>GLP-1 analogues and Glitins (dipeptidyl peptidase-4 (DPP-4) inhibitors)</td>
<td>• Increase glucose-induced insulin secretion and inhibit glucagon secretion (Drucker &amp; Nauck 2006, Cervera et al. 2008)</td>
<td>• Rare lactic acidosis side effects (Misbin 2004)</td>
</tr>
<tr>
<td>Alpha glucosidase inhibitors (AGI)</td>
<td>• Slow carbohydrate absorption in the gut to reduce post-prandial glucose and lipid levels (Buse et al. 2004)</td>
<td>• Risk of weight gain, congestive heart failure and fractures (Gegick &amp; Altheimer 2004)</td>
</tr>
<tr>
<td>SGLT 2 inhibitors</td>
<td>• Inhibit renal sodium-linked glucose transporter 2 (SGLT2) to decrease glucose reabsorption and increase urinary glucose excretion (Ferrannini et al. 2014)</td>
<td>• Only effective during initial stages of the disease when functional beta-cells are present (Hemmingsen et al. 2014)</td>
</tr>
<tr>
<td>Insulin</td>
<td>• Directly reduces glycaemia by increasing glucose uptake and inhibiting hepatic glucose production (Luzzi et al. 1988)</td>
<td>• Fail to provide adequate glycaemic control over the long term (Fu et al. 2012)</td>
</tr>
</tbody>
</table>

less side effects for T2D. Although many potentially useful drugs have been discovered, only 50% of patients with type 2 diabetes mellitus achieve adequate glycaemic control with currently available medications (Mitka 2013). This could be due to the inability of these medications to target the root causes of the disease. Therefore, to make a real impact on diabetes and to change the natural history of the disease, there is a need to discover new therapeutics that improve the overall function and survival of tissues and cells involved in T2D pathophysiology.

Moreover, as T2D is a result of alterations in multiple complex pathways, drug discovery focusing on a single biological target is often not effective. Therefore, a novel approach is necessary for the identification of new antidiabetic drugs that does not focus on a single specific target, but instead integrates the overall complexity of the disease.

Gene expression signature (GES): a powerful tool in the diagnosis and treatment of complex diseases

A GES is a small set of genes that have predictive power to differentiate the overall transcriptome in a cell or tissue in response to an external stimulus, without integrating the direct involvement of individual genes (Alizadeh et al. 2000, Konstantopoulos et al. 2011, Chibon 2013). In essence, it can be thought of as a genetic fingerprint of the overall state of the cell. The GES approach has been used extensively as a powerful tool in the diagnosis and treatment of complex diseases such as cancer. In general, to obtain a GES from a normal vs diseased state, the first step is to profile the global gene expression changes in both states using a high-throughput screening technique that measures the expression of a large number of genes simultaneously, such as next-generation sequencing.
Using multiple statistical methods, a minimal set of genes (typically <100) that are differentially expressed and are able to most accurately discriminate between normal vs diseased state is identified as the GES (Chang et al. 2011). This GES can then be measured in samples of interest using techniques such as next-generation sequencing, multiplex or standard real-time PCR. How this GES is used varies depending on the application as discussed below.

**Use of GESs in complex diseases**

Cancer, like many diseases, is a complex multifactorial cellular disease with high heterogeneity in gene expression and phenotype (Rabbani et al. 2014). The large number of genes associated with the growth and proliferation of cancerous cells makes the task of providing optimal treatment extremely difficult. Extensive variability exists in the survival and treatment of patients, and given the fact that most cancer treatments are cytotoxic in nature, choosing drugs for patients must be specific as some of the drugs are more toxic than others and the efficacy can vary among different patient subtypes. Therefore, for therapeutic purposes, it is important to understand how the individual patient’s disease state is different from the normal physiological state and what the outcome is likely to be if they receive a specific drug treatment (Kalia 2013).

A classic example for the success of the GES technique is the 70-gene prognosis signature developed by Veer and coworkers, which is currently on the European market and has the potential to predict patient survival and treatment outcomes for breast cancer (van ’t Veer et al. 2002). Several other studies also identified GESs for the treatment and diagnosis of cancer such as the 31-gene GES in colorectal cancer that showed predictive power to identify patients with high risk of recurrence (Wang et al. 2015b), the 44-gene GES from 112 oestrogen receptor–positive primary breast carcinomas that exhibited a high predictive power to predict the treatment outcome of patients with progressive disease to tamoxifen (Jansen et al. 2005) and the 41-gene GES derived from breast cancer stem cells that predicted the risk of metastasis and survival in breast cancer patients (Yin et al. 2014). Thus, studies have explored the GES method to identify and confirm gene sets to predict the response to treatment and to classify different tumours to predict responsiveness to different drugs (Sorlie et al. 2001, Ayers et al. 2004, Folgueira et al. 2005, Tabchy et al. 2010).

The use of GES has not been limited to cancer but has been used recently for infectious diseases such as tuberculosis (TB), malaria, influenza and acquired immune deficiency syndrome (AIDS). In particular, the study of TB can be complex as in most people infected with TB, the bacterium remains latent for a time with a high chance of developing disease later in life. Therefore, differentiating latent vs active infection states is important. Further, the immune response to TB is not well characterised making it difficult to develop new treatments, vaccines or even accurate diagnoses (Berry et al. 2010). Therefore, the prospect of a GES provided a promising path to provide more personalised treatment and diagnosis. Different studies applied GESs that differentiated healthy and diseased states as well as different stages of infection, such as the identification of the 393 transcript signature for active TB that is differentially expressed in whole blood of patients with active and latent TB vs healthy controls (Berry et al. 2010). Additionally, an 86-gene GES discriminates active TB from other inflammatory and infectious diseases (Berry et al. 2010). Moreover, another group has identified a GES that successfully differentiates TB patients, TB-infected healthy individuals and non-infected healthy individuals (Jacobsen et al. 2007). Thus, the GES technique may have an impact for vaccine and therapeutic development leading to a new era of protection in diseases such as TB.

The application of GES technology is not limited to the previously mentioned disease examples but is also valuable to study the adaptive immune response to vaccines to predict vaccine efficacy. One such promising work is the identification and use of a GES to predict immune responses to yellow fever vaccine YF-17D. The distinct GES predicted the magnitude of neutralising antibody response with up to 100% accuracy (Querec et al. 2009). This approach is now extensively used to study the innate and adaptive immune response to vaccination against various infections such as influenza, HIV and malaria. Categorising the immune response to infections is usually difficult but is crucial for treatment purposes. This could enable more efficient diagnostics as well as a better understanding of the pathogenesis of infection. An ‘acute respiratory viral GES’ was successfully identified, which was confirmed to have the ability to differentiate between healthy uninfected controls vs acute respiratory infected patients as well as viral vs bacterial infection (Zaas et al. 2009). It is now clear that GES-based techniques have potential to improve diagnosis and treatment outcomes for a range of complex diseases.
GES for drug discovery

Conventionally, drug discovery for complex diseases was carried out by screening small molecule libraries that bind or affect a specific target. However, in some cases, the mechanisms underlying the cause of these diseases are highly complex and usually involve the interaction of multiple signalling pathways (Evans & Guy 2004). Hence, compounds identified through conventional drug discovery methods frequently fail to modulate the underlying cause(s) of complex diseases. Unlike the traditional method, the high-throughput GES-based screening technique does not require prior knowledge of the target; instead this method will identify compounds based on their effect on the overall cellular state represented by the GES. This method has the potential to be a more powerful approach for drug screening for complex diseases such as diabetes.

The GES method uses a defined set of conditions, usually in a cellular model, to create at least two states of interest from which the transcriptome is measured and modelled to generate the GES. This same model can then be treated with a drug library and then the expression level of the GES genes measured using a technique such as qPCR to identify drugs that have a positive influence on the GES genes and are therefore likely having the desired effect in the cell model. By measuring a single and simple readout of qPCR rather than having to carry out more involved and numerous cell characterisation assays, this streamlines the drug discovery process. In addition, the readout measurement of qPCR can be easily performed in a 96- or 384-well format making it more high throughput.

As the GES method is an unbiased and high-throughput approach, many different drug libraries can be screened for efficacy. In fact, by screening available libraries of off-patent, FDA-approved drugs, the timeline from drug discovery to clinical use can be significantly truncated. This approach has been used previously to fast-track an insulin-sensitising drug from GES generation to a successful Phase 2 clinical trial in just three years (Konstantopoulos et al. 2011, Simpson et al. 2014) (discussed in detail in the following section).

Efficacy of the GES method for drug discovery was confirmed by Stegmaier and coworkers using primary acute myelogenous leukaemia (AML) cells (Stegmaier et al. 2004). Using these cells, they generated a GES that differentiated the gene expression pattern of untreated primary AML cells vs differentiated myeloid counterparts (neutrophils and monocytes). The GES was then used to identify candidate compounds that can induce differentiation of AML cells and successfully identify compounds that regulated the GES genes and caused a transition pattern that was similar to actual differentiation (Stegmaier et al. 2004).

In another study, Antipova and coworkers demonstrated the possibility of using a GES in drug discovery where they examined platelet-derived growth factor receptor (PDGFR) signalling, in particular, the extracellular regulated kinase (ERK) pathway, which is often upregulated in tumours and is a target for anticancer therapies (Antipova et al. 2008). They developed a GES that represented PDGFR activation and screened a compound library of small molecules to find compounds that would reverse or turn off the signature. Two compounds were identified that reversed the signature including one compound that was an inhibitor of phospholipase A2, a known regulator of ERK signalling, with the second compound being a novel discovery. These studies demonstrate the potential of using GES technology in drug discovery for complex diseases.

A limitation however to using GES for drug discovery is that a defined and robust set of experimental conditions, such as treated cell lines, needs to be available from which the GES can be generated. These same defined conditions will then be treated with compounds from a compound library and the GES measured to identify candidate drugs. Therefore, the quality of the GES, and thus, the drugs it can discover, is limited by the model used. This limitation is not a problem for drug discovery of well-defined diseases/conditions for which an appropriate model exists; however, for diseases for which we have little understanding or no appropriate cellular model, the approach of GES for drug discovery will be limited. In addition, development of a GES requires complex statistical computational methods making it essential to have bioinformatic expertise; however, the development of user-friendly modelling packages is making this less of a limitation (Li et al. 2013 #1243).

GES in T2D drug discovery

The first and only study to apply the GES approach to drug discovery in T2D was aimed at reducing insulin resistance. Konstantopoulos and coworkers used an adipocyte cell line (3T3-L1) treated with tumour necrosis factor-α (TNF-α) to induce insulin resistance. They then further treated these cells with the known insulin sensitising agents, aspirin and troglitazone to ‘re-sensitise’ the cells to insulin, which was their ‘biological state of interest’. After the successful
reversal of the insulin resistance state, transcriptome profiling was performed using microarray and a GES consisting of 5 genes was generated using Bayesian model selection to identify a small subset of genes that best represented the ‘re-sensitised state’. This GES was able to successfully discriminate between T2D individuals with low vs high insulin sensitivity showing that it can be used to efficiently subtype patients (Konstantopoulos et al. 2011). The GES was then also used to screen a compound library to identify pharmacological agents that demonstrated similar gene expression pattern to that of the insulin ‘re-sensitised state’ (Konstantopoulos et al. 2011). Out of the identified compounds, one, the carbonic anhydrase inhibitor methazolamide (MTZ), was then further investigated. MTZ exhibited glucose-lowering effects and enhanced glucose tolerance in animal models of T2D (Konstantopoulos et al. 2012) and successfully improved glycaemic control in T2D patients in a Phase 2 clinical trial (Simpson et al. 2014) indicating that this agent is a promising new agent to improve insulin sensitivity in T2D. Although this drug discovery endeavour was successful, the complex and multi-tissue nature of T2D means that targeting insulin resistance alone will not combat all the problems that occur in this disease. However, the technique of using a GES for drug discovery can be easily applied to other tissues that play a role in T2D pathophysiology to identify further drugs that can be used in combination to treat multiple aspects of this disease.

**Conclusion**

With diabetes affecting more than 422 million adults worldwide (NCD-RisC 2016), the need to discover and develop improved therapies is urgent. Traditional antidiabetic drug interventions are usually initially beneficial; however, as the disease progresses, these drugs eventually fail to improve the treatment outcomes. This failure is linked with the incompetence of these drugs to target the root cause of the disease as well as their inability to modulate the multiple complex pathways associated with the disease. Recent advances in genome-wide association studies provided considerable hope for greater understanding of the complexity of the disease; however, we are still a long way from understanding the pathophysiology of complex diseases to sufficiently identify appropriate targets. Further, even if we do identify possible targets, modulating on risk pathway is unlikely to have a beneficial overall effect. The GES-based drug discovery approach is an alternative method for drug discovery, where the drugs are selected based on a more complete measurement of how a cell is responding to them (Schadt et al. 2009) and has proved successful in other complex diseases. Therefore, we propose that GES-based drug discovery is likely to be a powerful tool in the field of complex diseases including T2D.

**Declaration of interest**

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

**Funding**

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

**Author contribution statement**

S S and K A-M conceived and drafted the manuscript and were responsible for all sections. T C and K W contributed to writing the manuscript and revised it critically for intellectual content. All authors approved the final version of the paper.

**References**


Aston-Mourney K, Proietto J, Morahan G & Andrikopoulos S 2008 Too much of a good thing: why it is bad to stimulate the beta cell to secrete insulin. Diabetologia 51 540–545. (doi:10.1007/s00125-008-0930-2)


Mishra RI 2004 The role of type 2 diabetes mellitus in the development of type 2 diabetes mellitus in patients with diabetes. Diabetes Care 27 1791–1795. (doi:10.2337/diaca.27.7.1791)

Mitka M 2013 More patients get good diabetes control, but only a minority meet all goals. JAMA 309 1335–1336. (doi:10.1001/ jama.2013.2414)


Motala AA, Omar MA & Gouws E 1993 High risk of progression to NIDDIM in South African Indians with impaired glucose tolerance. Diabetes 42 556–563. (doi:10.2337/diab.42.4.556)


Received in final form 28 November 2016
Accepted 7 December 2016
Accepted Preprint published online 7 December 2016