Léo Blondel, Ecole Normale Supérieure

Master 1 Internship report

# Transdifferentiation prediction using gene regulatory networks and genetic algorithm
# &
# Functional co-culture of motor neuron and myotube

## Abstract:

Cellular reprogramming, also called transdifferentiation is the process which change the fate of a defined cell type to another one. To reprogram a cell, researcher use a cocktail of defined transcription factors that induce changes in the molecular and genetic state of the cell. However they have to start with a big number of transcription factor and then reduce it. Using a motif based prediction algorithm and graph theory as well as a genetic algorithm, we constructed the gene regulatory network and scored factors to predict the importance of each of them in the reprogramming process.

Amyotrophic Lateral Sclerosis (ALS) is a motor neuron degenerative disease poorly understood. Motor neuron normally form Neuro Muscular Junction with myotube, and that junction seem to be implicated in ALS. Here we propose a new *in vitro* coculture system of human embryonic stem cell derived motor neuron and chicken myotubes. That system allows study of the Neuro Muscular Junction formation and might lead to a better understanding of ALS as well as drug screen. We also propose an *in vivo* model to study the behavior of motor neuron injected in the neural tube of a chicken embryo.

# Report Outline

# Introduction

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease of early onset (30-40 years old) in which motor neurons (MNs) degenerate from cell autonomous and non-cell autonomous signals (Dimos et al., 2008) leading to progressive paralysis and patient death from respiratory failure. Even though some mutations have been previously described (Chiò et al., 2009; Sha, Z. Zhang, Schymick, Traynor, & S. Zhang, 2009; Shatunov et al., 2010) and well studied, only 10% of the ALS patients carry them and most cases are sporadic. One of the mutations is carried on the SOD1 gene: G93A (SOD1G93A) (Karumbayaram et al., 2009) and a lot of studies have focused on that particular mutation. Previous co-culture experiments have shown deleterious effects of SOD1G93A glia cells on the survival of MNs (Dimos et al., 2008)(SAXEL, 1977). The next step in understanding the effect of that mutation on human MNs (hMNs) was to study the neuromuscular junction (NMJ). The classic cell line for myotube and NMJ *in vitro* study is the mouse myoblast cell line C2C12 (SAXEL, 1977) but previous experiments with C2C12 cells did not work well, resulting in a poor NMJ formation especially with hMNs. However, it has been shown that the use of primary chicken myoblasts harvested from chicken embryos and differentiated into myotubes resulted in mouse embryonic stem cell (mESC) derived MNs in a good NMJ formation (unpublished result). Therefore, we chose to use that system to setup the co-culture. Furthermore, an *in vivo* motor neuron survival assay, using chick embryo as a chassis for the motor neuron to develop, was setup through the microinjection of hMN differentiated embryoid bodies (EB) into the neural tube.

Another issue was to obtain human SOD1G93A hMNs. The use of patient specific iPS cell lines is a good approach, but another way to create MNs would be by transdifferentiation, also called cell reprogramming. Up until Yamanaka's iPS discovery (Takahashi & Yamanaka, 2006), cell reprogramming was considered as doable mainly by somatic cell nuclear transfer, an inefficient and complicated system. Since we have the ability to reprogram cells using simple retrovirus systems, a lot of effort has been put into it. The reprogramming process consists of turning one cell type into another, changing what we believe to be a fixed fate, and is called transdifferentiation. This process can occur naturally in a few organisms but remains rare and poorly understood. Over the past few years, scientists have succeeded in transdiferentiating cells *in vitro* using over expression of a specific subset of key regulatory genes. For example, Yamanaka's four factors, Oct4, Klf4, Sox2 and c-Myc, can reprogram a differentiated cell into a pluripotent one that is now called, an Induced Pluripotent Stem cell (iPS). But other trans-differentiations have been successfully performed, such as pancreatic cells into hepatocyte (C. N. Shen, Horb, Slack, & Tosh, 2003), fibroblast into neuron (Janghwan Kim et al., 2011; Vierbuchen et al., 2010) and recently into MNs (Eggan, K., unpublished result), fibroblast into cardiomyocyte (Efe et al., 2011), etc. All of those systems have in common the use of master key transcription factors (TFs). Through the expression of these factors, the cell undergoes drastic gene expression changes which eventually will change its fate. Even if some succeed in that process, the research principle consists mainly of brute forcing the cell. Indeed, the majority of those transdifferentiation projects start by a selection of potential TFs, usually by using microarray data, and then try different cocktails of genes. But this process is time consuming and expensive. In order to try to gain a better understanding of this process, we tried to setup a prediction algorithm. The idea was that through the study of the transcription network using a Gene Ontology (GO) scoring system, it would be possible to rank those genes for transdifferentiation. To do so, DNA TF binding predictions were used to construct the network. Later, microarray data were used to refine the prediction. And ultimately a genetic algorithm was used to predict the GO scoring system.

Here we show that a putative transcription network is predictable with good precision, but also that the use of microarray data in the network allows a refinement as well as a better understanding of the cell state. Finally, the use of a genetic algorithm is a good approach to find important functions that participate in the transdifferentiation.

# Materials and Methods:

## I. Human Motor Neuron functional assay

### a. Muscle extraction protocol

Fertilized Chicken eggs were incubated for 12 days at 37.5°C. The embryo was extracted by cutting through the egg shelf and transferring the egg content into a 15cm petri dish filled with PBS. Then they were exsanguinated by cutting the umbilical cord. The pectoral muscle was then harvested, transferred into 10 ml of DMEM + 10% FBS + 100ug/ml Penicillin/Streptomycin (Chick Media) triturated with a sterile razor blade then a 10ml pipette, and finally a P1000 micropipette. The cells were centrifuged 15 sec at 1000 RPM and the supernatant was transferred into a clean 15ml conical tube. The cells were then plated onto a Poly-D-Lysine/Laminin coated tissue culture dish at a density of 100 000 cells/$cm^2$. The media was changed every 3 days.

Previously reported culture systems were described using another media: Ham's F10 medium, 10% Horse Serum, 5% Chicken Serum and 125mM CaCl2, 100ug/ml Penicillin/Streptomycin.

### b. Cell Culture

All cells were cultured at 37°C and 5% $CO_2$.

#### *i. Human Embryonic Stem cells:*

Human Embryonic Stem cells were cultured in two different ways:

On Matrigel: HuES3 HB9::GFP (motor neuron reporter gene) were seeded on Matrigel (BD Bioscience) coated tissue culture dish and cultured in mTeSR1 (StemCell technologies). Media were replaced daily for the duration of hES expansion, and the cells in these conditions were passaged every 5-7 days using a solution of 1mg/ml Dispase (StemCell Technologies).

On feeder cells: HuES3 HB9::GFP were cultured as described by Cowan et al. (2004). They were maintained on a feeder layer of inactivated mouse embryonic fibroblasts (GlobalStem) in hES media (KO-DMEM (GIBCO), 20% KO Serum Replacement, 10,000 units penicillin, and 1mg/ml streptomycin (GIBCO), 2 mM glutamine (GIBCO), 100 µM nonessential amino acids (GIBCO), 55 µM beta-mercaptoethanol (GIBCO), 10 ng/ml bFGF2 (GIBCO)). Media were replaced daily for the duration of hES expansion, and the cells in these conditions were passaged every 5–7 days by 0.05% trypsin (GIBCO) gentle dissociation.

#### *ii. MNs directed differentiation: Human*

IPSCs and ESCs were differentiated as described previously (Boulting et al., 2011), but with the following modifications (Fig2a): differentiations were started from dispased colonies triturated to become ~50-cell aggregates of iPSCs, and from days 1–7 were cultured in the presence of RepSox (10 µM) and Dorsomorphin (0.2 µM, Stemgent) to neuralize the cultures. From day 5 onward, BDNF (10 ng/ml, R&D), ascorbic acid (0.4 µg/ml, Sigma) and Retinoic Acid (RA) (Sigma) were added.
From day 7 onward, Smoothened Agonist 1.3 (SAG) (Calbiochem) was added at 0.5 µM. EBs were dissociated, plated and assayed as described above on day 21.

#### *iii. Primary Glial Cultures*

P1–P3 mouse pups were sacrificed by using an approved method of euthanasia. Under a dissection microscope, the parenchyma were isolated and the meninges were carefully stripped away with fine forceps. The tissue was then dissected into small pieces and transferred to a solution containing 12 ml of HBSS, 1.5 ml of trypsin (GIBCO), and 1% DNase (Sigma) and incubated at 37°C for 15 min, swirling the mixture periodically. The supernatant containing the dissociated cells was collected, and 3 ml of FBS was added to inhibit the trypsin.

Cells were then centrifuged at 1000 rpm for 5 min, re-suspended in glia medium (minimum essential medium (MEM) with Earle's salts (GIBCO), D-(+)-Glucose 20% (Sigma), penicillin-streptomycin (GIBCO), 10% Horse Serum (GIBCO)) and plated at the concentration of 80,000 cells per ml in T75 flasks (Falcon). After the glia reached confluency, they were re-plated onto Poly-D-Lysine/Laminin coated tissue culture dish (BD Biosciences).

#### *iv. Human motor neuron culture*

Glial cells were harvested and cultured as described above, cells were then trypsinised using 0.05% trypsin and plated on Poly-D-Lysine laminin coated dish with glia media. 24-48h later, hMNs were seeded on top of it and the medium was switch to hMN media.

For the RepSox experiment, 10µM RepSox was added to the hMN media and the treatment started as soon as the hMN were plated.

### c. Dissociation of neuralized EB

EB at day 21 of neuralisation were dissociated using a 1% Papain solution with DNase. EB were washed with Phosphate buffered saline (PBS) twice, and then incubated for 30 min in the Papain mix. Cells were then gently triturated then either plated or sorted using flow cytometry.

### d. Co-culture assay

Myoblasts were extracted as described above and plated on Laminin coated 13 mm Plasma treated coverslips (Nunc) at a cell density of 100 000 cells/$cm^2$. Media was changed at day 3 and at day 6 of culture, neuralized hEB were plated on top of the muscle and the media was changed for hMNs media (Neurobasal (Invitrogen), NEAA, penicillin/streptomycin, heparin (2 µg/ml), N2 supplement (Invitrogen), B27 (Invitrogen), retinoic acid (RA) (1 µM, Sigma), ascorbic acid (0.4 µg/ml, Sigma), db-cAMP (1 µM, Sigma), 10 ng/ml of each of BDNF, GDNF and CNTF (R&D), 25 µM β-mercaptoethanol (Millipore) and 25 µM glutamic acid (Sigma)). 75% of the media was changed every other day.

### e. Injection

HuES3 HB9::GFP EB were neuralized as described above. Chicken embryos were incubated at 37.5°C for 50h, then hMNs were injected in the embryo as described previously(Boulland, Halasi, Kasumacic, & Glover, 2010), but with the following modifications: instead of removing one somite we removed for a length of 3 somites half of the neural tube at different locations (6-9th, 10-13th, 16-19th, 20-23th somites). Instead of using a cell suspension we micro-injected a single intact EB into the wound. The eggs were then incubated for 3 more days.

### f. Immunocytochemistry.

Cell cultures were fixed in 4% PFA for 15–30 min at 4 °C, permeabilized and quenched with 0.1–0.2% Triton-X in PBS (wash buffer) and 100 mM glycine (Sigma) for 20 min.

Cells were blocked in wash with 10% Horse Serum for 30 min and then incubated with rabbit polyclonal Anti-GFP-alexa-488 antibody (Invitrogen, 1:400) and α-Bungarotoxin-Alexa-594 (Invitrogen, 1:100) overnight. Images were acquired on a confocal microscope (Zeiss LSM510).

### g. Whole mount immunocytochemistry

A modification of (Klymkowsky & Hanken, 1991) whole-mount G. Kardon immunocytochemical technique was used to stain the hMNs in whole normal chicken embryo. Injected day 5 embryos were fixed overnight at 4°C with 4% paraformaldehyde and then bleached overnight with Dent's bleach (50% methanol, 10% DMSO, 15% H2O2). Specimens were washed for 3 hours in PBS, stained for 72h at 4°C with a rabbit polyclonal Anti-GFP-alexa-488 antibody (1:400) (in 10% serum, 20% DMSO), and washed for 24 hours in PBS at 4°C. Images were acquired with an epifluorescent dissection microscope (Leica).

## II. Transdifferentiation prediction using gene regulatory networks and genetic algorithm

### a. Database: NCBI, Jaspar, Transfac Pro

For this project, multiple databases have been used: the National Center for Biotechnology Information (NCBI), RefSeq database, as well as the annotation database and the Mouse genome upstream sequences. For the position weighted matrix (PWM) database, the Jaspar database (Bryne et al., 2008; Wasserman & Sandelin, 2004) as well as the Transfac Pro database (Matys et al., 2006) were used. The PWM from the Transfac Pro database were converted into the MEME format and merged with the Jaspar database which we will refer here to as the PWM database. For the functional annotations, the Gene Ontology Database (Ashburner et al., 2000) was used.

### b. Binding site prediction

The binding site prediction of a given TF was performed using the "fimo" software from the MEME Suite (Bailey & Elkan, 1994).A selected subset of PWMs were then ran against the 5kb upstream sequence of all genes in the mouse genome. The results were then refined with a p-value selection of $<1.10^{-7}$ as well as a fimo Score selection >15.

### c. Microarray

Microarray data were kindly given by Justin Ichida and were acquired using the Illumina technology.

### d. Transcription factor selection

To determine whether or not a selected gene is a TF, GO IDs were used. Using the NCBI annotation database and a selected subset of GO IDs, a list of TFs was extracted.

### e. Recursive analysis

In order to construct the transcriptional network of a subset of genes, a recursive fimo analysis was performed (using the fimo software from the MEME suite (Wasserman & Sandelin, 2004)). Starting with that subset, a fimo analysis was performed (M&M: Binding site prediction). Among the results, the TFs were extracted using the TF GO ID list and the binding site (BS) for each TF were stored. The new list of TFs was then reused as a new entry for the analysis. At each step the BS information are stored for each TF.

After 3 rounds of analysis, the network was constructed from the BS data following these rules: The network is directed, the edges are weighted, a node is a gene, an edge is a predicted binding site directed from the TF to the bound gene. Each edge has a weight ($\in [0;1]$) which is proportional to the quality of the fimo prediction following that equation:

$$Weight = \frac{p\ value - Minimum\ p\ value\ in\ the\ network}{2(Maximum\ p\ value\ in\ the\ network - Minimum\ p\ value\ in\ the\ network)} + \frac{fScore - Minimum\ fScore\ in\ the\ network}{2(Maximum\ fScore\ in\ the\ network - Minimum\ fScore\ in\ the\ network)}$$

In graph theory, every node and edge can contain an unlimited amount of information. A condensed version network was then used where only TF were assigned as node, and each node contained as attributes all the predicted genes as well as all the GO IDs of those genes.

### f. Microarray data refinement

The network previously created was then refined with microarray data. Each node was tested for its expression in a given system, if the gene wasn't upregulated (>20 fold) or downregulated (<-20 fold) then it was removed from the network.

Moreover, using the expression data, a possible interaction between two TF was assigned following the chart below were TF1 binds to TF2:

| TF1 | TF2 | Interaction |
|---|---|---|
| Upregulated | Upregulated | Activator |
| Downregulated | Downregulated | Activator |
| Downregulated | Upregulated | Repressor |
| Upregulated | Downregulated | Repressor |

### g. Genetic algorithm

The genetic algorithm was created following the rules of (Goldberg, 1989). A gene is a GO ID score, the genome is all the GO IDs in the network, the mutation rate is 0.1%, with a mutation intensity of +/- 10%. Every individual has a lifetime of 3 generations. After those 3 generations, its chances to die are proportional to its fitness.

# Result

## I. Human Motor Neuron functional assay

### a. Chicken myoblast extraction and differentiation into myotube.

In order to setup a co-culture assay, we first needed to extract a pure myoblast population and to differentiate them into mature, contractile, myotubes. It has previously been reported that harvesting the pectoralis muscle at stage 38 (day 12) of the chicken embryo followed by a muscle trituration leads to the extraction of a myoblast population which fuses over time into mature myotube (Rafuse & Landmesser, 1996). In that extraction process, cells were plated with a special media containing 5% of Chicken serum (CS) as well as 10% of Horse serum (HS) (M&M: Muscle extraction protocol). When we used that protocol, we did see some myotube formation, but the extraction process usually lead to an undifferentiated layer of myoblast and fibroblast (Fig1a).

We first tried to improve the extraction yield by harvesting more selectively the muscle, but we didn't see any improvement (Fig1b). Thus, we tried to improve the extraction yield by applying a selective pressure on the extracted cell population. To do so, we removed one by one all of the different elements of the media composition and tried to replace them with similar elements. We showed that replacing Ham F10 by DMEM medium, as well as HS and CS by 10% FBS, increases the myotube formation by over fivefold (Fig1c). We also showed, that the use of CaCl2 was unnecessary for the myotube formation, but increased the contraction frequency (Fig1d). To conclude, we demonstrated that the previously reported protocol might not be the most suitable to achieve a pure and fully differentiated population of myotube.

### b. Human Motor Neuron differentiation and Neuromuscular Junction

Next, in order to perform the co-culture we needed to differentiate ES cells into hMNs. We used a modified version of the previously described neuralisation protocol (Boulting et al., 2011) (M&M: Cell Culture) (Fig2a). We started with a genetically engineered Human ES cell line carrying the reporter gene GFP under the control of the HB9 promoter region (HuES3 HB9::GFP). HB9 (also known as Mnx1) is a specific gene expressed primarily in hMNs among the spinal cord neuronal population. We first dissociated the EBs (M&M: Dissociation of neuralized EB), sorted high GFP expressing HB9::GFP cells using flow cytometry, and plated them alone or on top of a day 6 myoblast culture (Fig2b and 2c respectively). However, the survival without glial cells wasn't long enough for the hMNs to form a NMJ. Thus, we decided to plate whole EBs on the myotubes (Fig 2d) (M&M: Co-culture assay). It allowed us to keep the hMNs long enough for the NMJ to form. In order to see the NMJ, we used a labeled protein, α-Bungarotoxin coupled with Alexa-594 (α-Bung). α-Bung is a neurotoxic protein first isolated in the venom of Kraits (an indian snake) which binds specifically to nicotinic and acetylcholine (ACh) receptor. It is well known that the α-Bung binds specifically to the ACh receptor of the NMJ allowing us to visualize that region (Rafuse & Landmesser, 1996). Using the double staining HB9::GFP and α-Bung-Alexa-594, it was possible to see the connection between hMNs and myotubes (M&M: Immunocytochemistry). After 17 days of culture some of the hMNs had successfully formed a NMJ with a chicken myotube (Fig2e, due to light diffraction of the Thermanox coverslips, we couldn't get a better picture). Here, we successfully co-culture hMNs with chicken myotube and we have been able to see the formation of NMJ between those two populations. Next, an electrophysiologic study of that junction will be necessary to make conclusions as to its functionality.

### c. TGF-β inhibition promotes motor neuron survival as well as myotube formation

The neuralisation protocol employs the use of a TGF-β inhibitor SB 431542 (M&M: Cell culture). We then tested another compound named RepSox (or E-616452) also known to be a TGF-β inhibitor (Fig3a). It was first discovered as a TGF-β inhibitor (Gellibert et al., 2004) and later described as a chemical replacement for Sox2 during iPS reprogramming (Ichida et al., 2009). We found that the neuralisation using 10µM of RepSox was as efficient as with the SB 431542 (unpublished result, efficiency of about 30%) (Fig3b). Our hypothesis was that if TGF-β signaling inhibition participates in the differentiation into hMNs then it might also have an effect on the mature neuron. Our first hypothesis was that it would help the survival, we therefore plated the hMNs with glial cells, with or without 10µM RepSox (M&M: Cell culture).

Our result showed a slight increase in the survival when the hMNs were plated alone (data not shown), and a stronger effect when they were cultured along with glial cell (Fig3c).

Moreover, TGF-β signaling has been previously described as a strong inhibitor of the myotube differentiation (Ge et al., 2011). We therefore tried to culture the myoblast with different concentrations of RepSox. We saw a clear dose response in the amount and the maturity of the myotube as soon as 3 days of culture and an even stronger effect after 5 days with 10µM as the best concentration (Fig3d). Knowing that the TGF-β inhibition will help the myotube formation and that the optimal concentration was the same as the one used for the hMNs, we tried to co-culture those cells with 10µM of RepSox. As expected we saw an increase in the survival of the hMNs but more interestingly we saw a strong effect on the axonal growth activity (Fig3e). Indeed, when we compared the EBs cultivated with RepSox against those without TGF-β inhibition we clearly saw that almost all the hMNs of the RepSox EB projected axons while in the control only a few hMNs projected axons (Fig3e). To conclude, we showed that inhibition of TGF-β among the hMN has a drastic effect on the axonal activity and a non-cell autonomous survival increase probably due to glial cells.

### d. hMN implantation in a chicken embryo

It is clear that an *in vitro* protocol to study the NMJ is a really useful tool as it allows for drug screens, but cells don't behave *in vitro* as they do *in vivo*, this is why we decided to implant an EB into a chick embryo at stage HH16-18 (50h of incubation). We used HuES3 HB9::GFP neuralized EB and we performed the injection as described previously (Boulland et al., 2010) (M&M: Injection). We harvested the embryo after 3 more days of incubation and we were able to see that some EB attached and started to send out axons in the embryo (Fig3f). We showed that it is possible to inject hMNs into the neural tube of a chicken embryo at day 2 and that the cells survived at least 3 days and start to form axons in the embryo. An interesting follow up would be to see if the cells successfully formed a NMJ therefore allowing us to study it *in vivo*.

## II. Transdifferentiation prediction using gene regulatory networks and genetic algorithm

It has been shown that by using a cocktail of 7 factors it is possible to transdifferentiate fibroblast into MNs (Eggan, K., unpublished result) called Induced Motor Neuron (iMNs). However, the underlying mechanism that allows those cells to change their fate remains largely unknown. It is usually the same for the actual reprogramming studies, indeed the process which allows a cell to change its fate, though well studied, remains unknown. Our hypothesis was knowing that functional gene annotations can be mainly summarized through the Gene Ontology (GO) database, if we postulate that most of the genes have annotations, then it would be possible to predict the trans-differentiation genes using a GO ID scoring system.

At present, the methods to differentiate a cell through a set of defined factors, starts by performing a microarray analysis between two cell types (usually fibroblast against the desired cell type), then by extracting the differentially overexpressed TFs and by scanning the literature for each one to select around 20 candidates. After doing so, a cocktail of the 20 genes is injected, and if it works, one would restrict the cocktail by removing one factor at a time. That technique allows the researcher to assess the importance of each factor in the reprogramation, and to restrict their cocktail to a subset of genes. But this process is long, expensive and requires a lot of manipulations. Thus, to reduce the time and the number of experiments we tried to predict *in silico* which factors are the most important for the transdifferentiation of one cell type to another (Fig5a). However, the first step, microarray analysis, was still necessary.

Indeed, this step allows us to select a small number of factors which are the most probable to be involved as they are overexpressed (Fig4a). Starting with the list of potential factors, we needed their gene regulation networks to understand by which mechanism they could act in the cell reprogramming. Some of the most studied pathways are already available, but most of them are still unknown, underlying the need for a method to create such a network.

### a. Transcription factor binding motif based analysis

TFs have the property to bind to a specific sequence of DNA, allowing the researcher to assign for each TF a probable DNA binding sequence (BS). There are a lot of ways to find that sequence, but each of them end by defining a Position Weighted Matrix (PWM) which gives the probability for each position of the sequence to be one of the 4 DNA nucleotides (Fig4b). Using those PWMs, it is then possible to predict the probability of a defined factor to bind to a DNA sequence. If that sequence contains a DNA motif close to the PWM, by doing a motif based research, it is possible to find probable binding sites for a given TF. There exist different databases of PWMs, but the most advanced ones are the private Transfac Pro database and the public Jaspar database. Thus, we decided to merge those two databases (M&M: Databases).

Since 2002, the mouse genome has been fully available (Waterston et al., 2002), but computing a motif based analysis on the 2.5Gbases of this genome requires a large number of computing resources. Also, even with the ability to predict the BS, we still need to correlate those BS with a putative regulated gene. Therefore, instead of using the whole genome we decided to use a subset of it. First to correlate a sequence with a gene, it is necessary to know its position, but thanks to the RefSeq database, each gene has an assigned position on the genome. Using the RefSeq database and the mouse genome, it is then possible to extract the 5 Kilobase (Kb) upstream sequence for every gene in the genome. Such databases already exist and are available for example on the Harvard website (“Mouse genome upstream sequences,” n d). We decided to use 5Kb because of the complexity of the gene regulation among eukaryotes. Indeed, gene regulation can act far from the transcription starting site pinpointing the necessity for such a big upstream sequence.

To perform a motif based analysis of PWMs on a DNA sequence two main tools are available, the "Match algorithm" from Transfac (Matys et al., 2006) and the "fimo algorithm" from the MEME suite (Wasserman & Sandelin, 2004). We decided to use the fimo tool to perform our analysis.

### b. Network construction

In order to construct a gene regulatory network we had to choose between constructing it for each TF in our subset or constructing the combined network of all the TFs together. Knowing the complexity of the gene regulation in eukaryotic organisms, analyzing each factor at a time would have introduced a bias, indeed those factors usually interact between each other with regulatory cycles, switches, etc. But computing all the factors together also increases the level of complexity of the network increasing the intrinsic connectivity. Nevertheless, to construct the network we used a recursive fimo analysis starting with all the TFs that we selected from the microarray (Fig5b) (M&M: Binding site prediction). We extracted all the probable BS for each of those factors, and then extracted the TFs from those putative regulated genes.

To extract the TFs, we use an annotation selection using GO IDs (M&M: Transcription factor selection).

We then used those TFs as our starting pool of factors and restarted the algorithm (M&M: Recursive analysis). Because the significance of each prediction decreases at each iteration, we arbitrarily decided to stop after 3 iterations.

Next, we needed a tool to analyze those interactions. Because of the power and the flexibility of graph theory, we decided to model our interactions using a graph (Fig5c). We defined a node as a gene, and an edge as an interaction. The flexibility of Graph theory allowed us to assign attributes to each node or edge. We then assign to each node their differential expression from the microarray as well as their GO IDs. To each edge were assigned the p-value and the fimo score as well as a weight. We defined the weight of each edge as a relative weight in the network (M&M: recursive analysis). Because of our interest in understanding the iMNs regulation network, we first created it using our algorithm (Fig4c). We needed a rough confirmation that our prediction was good, thus we decided to use the microarray data as a validation. Indeed our network was supposed to be the core network of MNs in order to reprogram them so every gene in our network should be either upregulated or downregulated. Defining an upregulation as a 20 fold increase and a downregulation as a 20 fold decrease in expression versus the fibroblast we showed that 89% of the genes in our network were affected in the cell, therefore validating our model.

### c. Refining the prediction using Microarray data

Our predictions were good but in order to refine them, we first started with a 2-dimensional selection on the p-value and the fimo score (Fig4d). We defined the false positive as a gene which is not differentially expressed and false negative as the non-prediction of a differentially expressed gene. We constructed the network using a selection of a p-value inferior at $10^{-5}$ and a fimo score superior at 10, and we then removed each node by decreasing the p-value and increasing the fimo score. At each step we tested the percentage of false positive and false negative, and we found that the best prediction score was a p-value of $2.10^{-7}$ and a fimo score of 18 (Fig4d). As we couldn't achieve a perfect removal of false positive without affecting the false negative using this approach, we also refined it by removing the genes which are not differentially expressed as they probably don't participate in the transdifferentiation process (Fig5c) (M&M: Using the Microarray data). Combining those two approaches we achieve an excellent quality of prediction.

### d. Scoring the factors

Now that we had our gene regulation network, to predict the importance of each factor we needed to score each factor according to their importance in the transdifferentiation process. The major issue was that this process is poorly understood, and it is difficult to assess the functionality of each gene from a gene regulation network because of its intrinsic complexity. To solve this problem we decided to implement two methods, the first one is directly inspired from the classical transdifferentiation experiment and consists in removing one factor at a time and measuring the disturbance in the network. The second way consists of scoring each TF using their GO IDs and the GO IDs of the gene they activate (Fig5e). Indeed, each method taken separately would give a result which is biased by multiple factors, like the quality of the prediction or the number of genes in a subnetwork, while using both simultaneously will reduce those.

To measure the disturbance induced by the removal of one factor at a time, we first created the subnetwork for each factor, then we removed that subnetwork from the main network simulating the removal of one factor.

To assess the disturbance induced by the removal of that subnetwork, we compared the number of nodes, the number of cycles, the number of self-loops, as well as the total GO ID score of the remaining network. Those numbers were then normalized to give a score between 0 and 1.

To measure the score of each factor, we made two hypothesis, first that each jump from one node to another decreases the importance of the next node, and secondly that each node can be scored according to its GO IDs, which represent its cellular functions, enzymatic reactions, and cellular compartment. The first issue was that one node is connected to a set of other nodes, some of those nodes are also connected to other nodes, and if you have a cycle, then you can't score any node of the cycle as they depend on the other nodes. Thus, we decided to use another scoring method, using Dijkstra's algorithm (Dijkstra, 1959) we determined the shortest path from each starting node to all the other reachable nodes in the network reducing the network to a Tree (Fig5e: GO ID scoring method). Then we scored each target node by taking its own score, and by multiplying it by its weighted coefficient. The coefficient was calculated by taking the weight of all the edges between the first node and the targeted node and by multiplying them between each other. Because the weight is inferior or equal to 1, at each jump the coefficient was decreased, thus reducing the importance of each gene with its distance and the quality of its prediction.

### e. Use of a genetic algorithm to create a scoring table

Now that we had everything ready to score the network, we needed a scoring table which assigns for each GO ID a score. Because the transdifferentiation mechanism is poorly defined, we couldn't build it by hand. Moreover, if we did, we would introduce a human bias. Thus we decided to use a genetic algorithm to setup a GO ID score table. But we needed at least two different networks from two different experiments, otherwise the selection process would have selected a certain number of GO IDs which are not transdifferentiation specific but cell-type and experiment specific. Thus, we decided to use the iPS Network as an external network, because it is well defined and studied. We defined a "gene" as a GO ID score and an individual as the set of all genes (an individual = a scoring table which assigns a score for each GO ID), and extracted all the GO IDs from both networks to create the first individual. We decided to use a classic genetic algorithm method (Goldberg, 1989) (Fig5d) consisting of multiple steps: An initialization which generates a population by randomization of the score for each gene and an attribution of a fitness, then a mating between two individuals selected according to their fitness (the higher the fitness the higher the probability of mating). The mating consists of random crossovers between the two parents which generate two children. Those two are then mutated, and, by a selection process (Fig5e) assigned a fitness.

The critical part of a genetic algorithm is the selection process. We decided to use our knowledge of the iMNs and iPS transdifferentiation as a goal to reach for the individuals. We constructed their networks starting with 20 selected factors, and we assigned a score to each factor, 1 if the factor was necessary, and 0 if it was unnecessary based on experimental data. Then, we calculated for each individual their disturbance score, as well as their GO ID score. Each score was then normalized for each individual, by taking the maximum and minimum score of the set of factors and to normalize each factor with those.

We then did the mean of the disturbance score and the GO ID score, and we defined the fitness as the correlation coefficient between the individual score and the known score (Fig5e).

By running the selection process on both the iPS and iMNs network and by doing the mean of the two fitness, we hypothesized that it would reduce the non-transdifferentiation noise. Moreover, every time that a transdifferentiation set of gene is discovered it is possible to add it to the algorithm and each added network should reduce the non specific noise.

Sadly because of a lack of time we could not run the algorithm long enough to select the GO IDs, and so we couldn't achieve a GO ID score table. Indeed, we managed to increase the speed of the selection by reducing to a minimum the number of calculations at each step but the time for the selection of each individual was still 0.2 +/- 0.17 making it hard to compute over a big number of generations. However, we gained confidence that this method could work, allowing researchers to reduce the number of factors they have to test, as well as revealing some mysteries of the transdifferentiation process.

# Discussion

## I. Human Motor Neuron functional assay

In order to culture chicken myotube we started by improving upon the extraction protocol (Rafuse & Landmesser, 1996). To do so we tried to apply a selective pressure which will only select the muscle cells. By reducing the total serum percentage from 15% to 10% as well as using DMEM and FBS instead of Ham's F10 and HS/CS, we've been able to selectively grow only the myoblast population. We also showed that the use of calcium was not necessary but helped the maturation of the myotube. Indeed to contract myotubes need to store Calcium ions in their endoplasmic reticulum, increasing the contraction activity. Reducing the serum percentage and using DMEM, reduces drastically the amount of fibroblast in our culture and avoided an overgrowth of that population, allowing the myoblast to fuse into myotube. Indeed it is known that overgrowth of fibroblast is one of the major issues in myoblast extraction (Pacak & Cowan, 2009). So our protocol did not directly improve the growth of the myoblast, but instead reduced the fibroblast growth. Indeed, using our protocol, we saw the first mature myotubes after 6 days of culture, while with the old protocol the first mature myotubes appear after 3 days of culture. But, because we wanted the purest population of myotubes in order to co-culture them with the MNs we decided that this protocol was the best one in our particular case. Also, our protocol is more reliable as it always gives rise to a good myotube population while the other one depends mainly on the quality of the dissection.

Using this myotube population, we successfully plated and cultured differentiated hMNs on top of it. We first had to find a way to culture those cells without glia as the survival wasn't long enough. We had two options, try to co-culture 3 types of cells, or using the non specific differentiation of the EB as a potential source of glial cell. The first solution didn't seem a good idea because the more cell types you try to co-culture, the more issues you can encounter. Thus, we plated whole EBs containing hMNs as well as a lot of other cell types on top of our muscle. It was our hypothesis that by doing so, other cells in the EB will help the survival of the neuron, and that's indeed what we saw.

Our main question was to know if the hMNs were able to bind to the chick myotube, we then assessed that question using a double staining, GFP, and α-Bungarotoxin-Alexa-594, allowing us to see the NMJ, as well as the axon. We've been able to see that some hMNs successfully bound to the myotube.

The main goal of this co-culture experiment was to use it to assess if there is any difference between ALS patient derived MNs versus wild type MNs. But we've been unable to differentiate patient specific MNs from the iPS cell line at this time.

However, we showed that the inhibition of the TGF-β pathway had a drastic effect on the MNs. Indeed, the use of a TGF-β inhibitor in the differentiation protocol made us think that inhibiting this pathway might have an effect on the MNs. We showed that the inhibition of TGF-β induced a slight increase (3-fold) in the survival of the MNs but mainly increased the axonal activity. Interestingly, the increase in the survival was only seen when the MNs were cultured with glial cells which lead us to think that the survival is a non cell autonomous effect, probably glia dependent. Interestingly we also saw that the increase in axonal activity wasn't glia dependent, and thus probably a cell autonomous effect.

As it is well known that TGF-β inhibition increase the myoblast differentiation into myotube, we tested RepSox with different concentrations on our myoblast. We indeed saw a big increase in myoblast fusion as well as in the contractile activity. Because RepSox was able to increase the survival of hMNs, we decided to test its effect on the NMJ. The main issue that we had was that inhibiting TGF-β will have an effect not only on the MNs but also on the muscle cell, making it difficult to assess if the observed effect was due to the myotube or the MNs. A better experiment would be to culture the muscle and the MNs in two separated chambers with an axon permeable membrane between the two chambers. As the axon would be able to go from one chamber to another, small molecules like RepSox would probably also be able to cross, but we believe that it would at least reduce its effect on the other cell type. Next, it would be interesting to better study the effect of RepSox on the hMNs as this could lead to a possible cure for ALS patients.

Finally, we successfully injected hMNs in the neural tube of a day 2 (Stage HH16-18) chicken embryo and we've been able to see some axon formation. This experiment is really interesting as it offers an *in vivo* model to study hMNs. Indeed, in-ovo experiments on chicken embryos (HH xxxx stage) are more simple than similar experiments on murine embryos at a similar stage of development (E9.5-E10.5) which would involve in-utero surgery. Chicken are in that case an easier model and we show here that it allows for the survival of hMNs as well as axonal activity.

## II. Transdifferentiation prediction algorithm

Here we explored a new approach to understand the mechanism of transdifferentiation and a new way to try to predict the importance of each factor during the transdifferentiation process. To build the gene regulatory network, we decided to use a prediction method using PWMs and the mouse genome, as well as a motif based research. Such an approach allowed us to predict a fairly good regulatory network. The choice to start with all the factors appears to be a better choice as it allowed us to study some of the interaction between genes. For example the number of cycles is increased in the main network versus the sum of the small networks. Moreover, use of Graph theory to construct the network appears to be the best choice as it allowed us to easily study it.

Indeed, not only have we been able to use some of the properties for our prediction, but graphs allows us to go further into the study of regulatory networks and transdifferentiation.

Nevertheless, we also showed that even if a motif based prediction was already a good prediction, the use of microarray data was necessary to better refine the network.

Indeed, without refinement, the network predicted 89% of the differentially expressed genes, while with a microarray refinement we've been able to achieve 94.6%. We've been unable to achieve 100% as the microarray data didn't contain the expression level of every gene and some measurements were statistically non significant (p-value > 0.01).

However, our approach missed an epigenetic analysis as the transdifferentiation process occurs not only with protein-DNA interaction but also with a lot of changes in the methylation state as well as changes in the histone code, leading to condensation/release of the DNA in the cell. Next we would like to integrate DNA methylation and epigenetic changes in the prediction. Also, in eukaryotes the process of regulation of genes usually doesn't involve only one gene but multiple protein interactions at the transcription site, as it has been demonstrated to be essential for example in the ES/iPS cell pluripotency network (Orkin et al., 2008). Thus, a study of interactions between elements of the network and patterns would allow us to better predict the expression of genes.

Thanks to the effort of the community of scientists around the world, the Gene Ontology database gains information every day. It is our belief that over time the use of such annotation will be crucial to the understanding of biology. Indeed, the level of complexity of a cell is so high that to understand part of it, annotations are critical. Here we used this database to try to predict functionalities which are critical for the transdifferentiation process. Even if we've been unable to complete the analysis for the transdifferentiation process, our refined network associated with the Gene Ontology is a powerful tool to understand mechanisms in the cell. Indeed, it is possible to cluster genes with their GO IDs, extract pathways, or target more specific cell functions.

However, in order to predict the trans-differentiation through the network we needed a scoring table. Our approach through a genetic algorithm seems to be the best idea. Indeed, we could select by hand a subset of GO IDs that match our current view of the transdifferentiation. But our understanding of this process is still incomplete and by putting such a bias we might have missed some new properties that were unexpected. Also, our current knowledge might be too small to correctly predict the important factors, underlying the importance of such an unbiased approach.

Finally, because we tried to remain as flexible as possible, our approach allows the use of new transdifferentiation networks to increase its accuracy. Indeed, our genetic algorithm is capable of computing the scoring table against a lot of networks. However the downside of such an approach is that each other network will increase the calculation time. But computers are more powerful every day and such computation shouldn't be an issue.

## Acknowledgment

I want to thank all the people from the Eggan Lab for their support throughout my internship and particularly Kevin Eggan for accepting my application and our rewarding conversation, Justin Ichida and Esther Y. Son for their amazing help and Nick Atwater for the HuES culture. Also Michael Ziller from the Meisner group who helped me with the bioinformatic analysis and for letting me use their Transfac Pro database license. Finally, I want to thank Mary Paschaki and Jose Riviera Feliciano for \their help with the chicken embryos experiment as well as Julien Bouissac for his help on this report.

## Figures

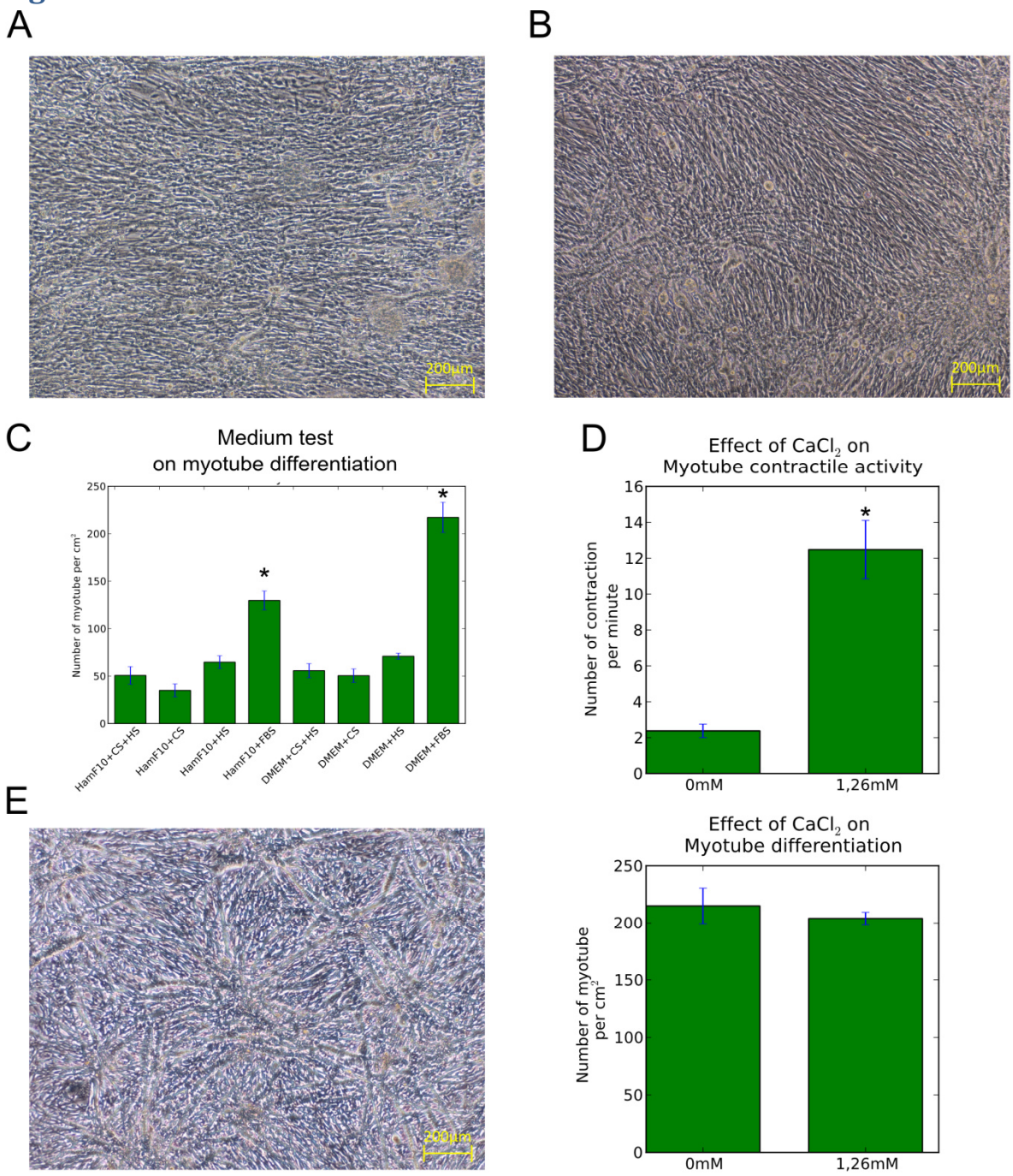



Figure 1: Myoblasts extraction and differentiation into myotubes.
a) Myoblast and fibroblast layer at day 7 after extraction. Few myotubes start to form but the layer is mainly undifferentiated.
b) Myoblast and fibroblast layer at day 7 after extraction. The myoblasts are aligned, but they don't fuse into myotubes.
c) Chart of the Medium selection. 4 well per condition, a Student t-test was performed. We can see here that the myotube formation is greatly improved under the FBS condition, and particularly when used with DMEM. (* : p-value < 0.01).
d) Chart representing the necessity of calcium. We counted the number of myotube as well as the number of contractions per minute. 3 wells per condition, a Student t-test was performed. (* : p-value < 0.01).
e) Differentiated Myotube grown with DMEM + 10% FBS + 1.26mM CaCl2. The layer is mainly differentiated, and fully contractile.

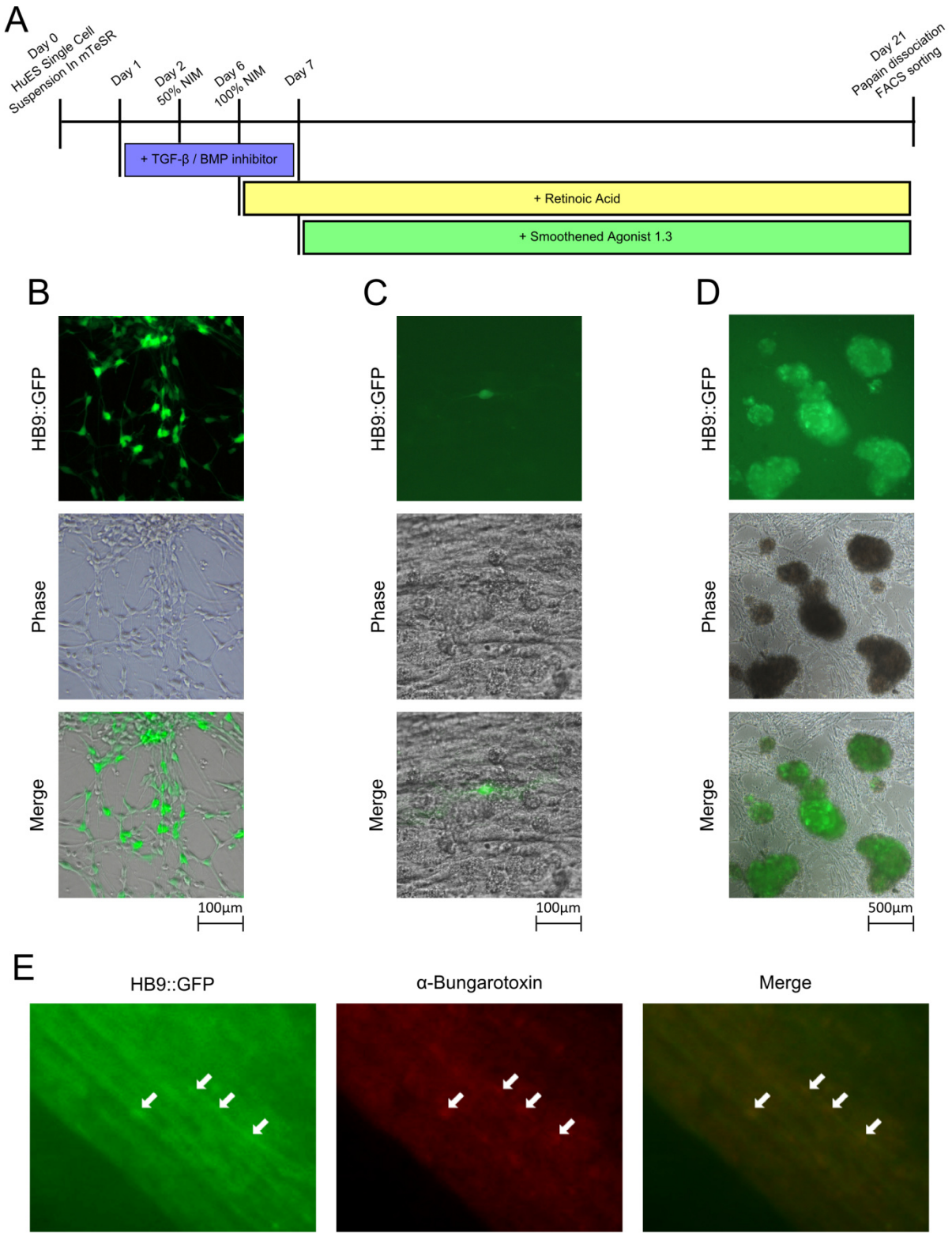



Figure 2: Motor Neuron Differentiation and myotubes co-culture
a) Scheme of the neuralisation protocol.
b) Day 7, HuES3 HB9::GFP derived MNs plated on Poly-D-Lysin Laminin coated tissue culture dish.
c) Day 5, HuES3 HB9::GFP derived MNs plated on top of 7 days old myotubes.
d) Day 3, HuES3 HB9::GFP neuralized embryoid body plated on top of 7 days old myotubes. The MNs start to spread in the dish 2 days after, and 3 days after it is already possible to see single MN.
e) Confocal imaging of the Neuro Muscular Junction between chick myotubes and hES derived MNs. Here we can see that neurons project axons along the muscle fibers and form NMJ.

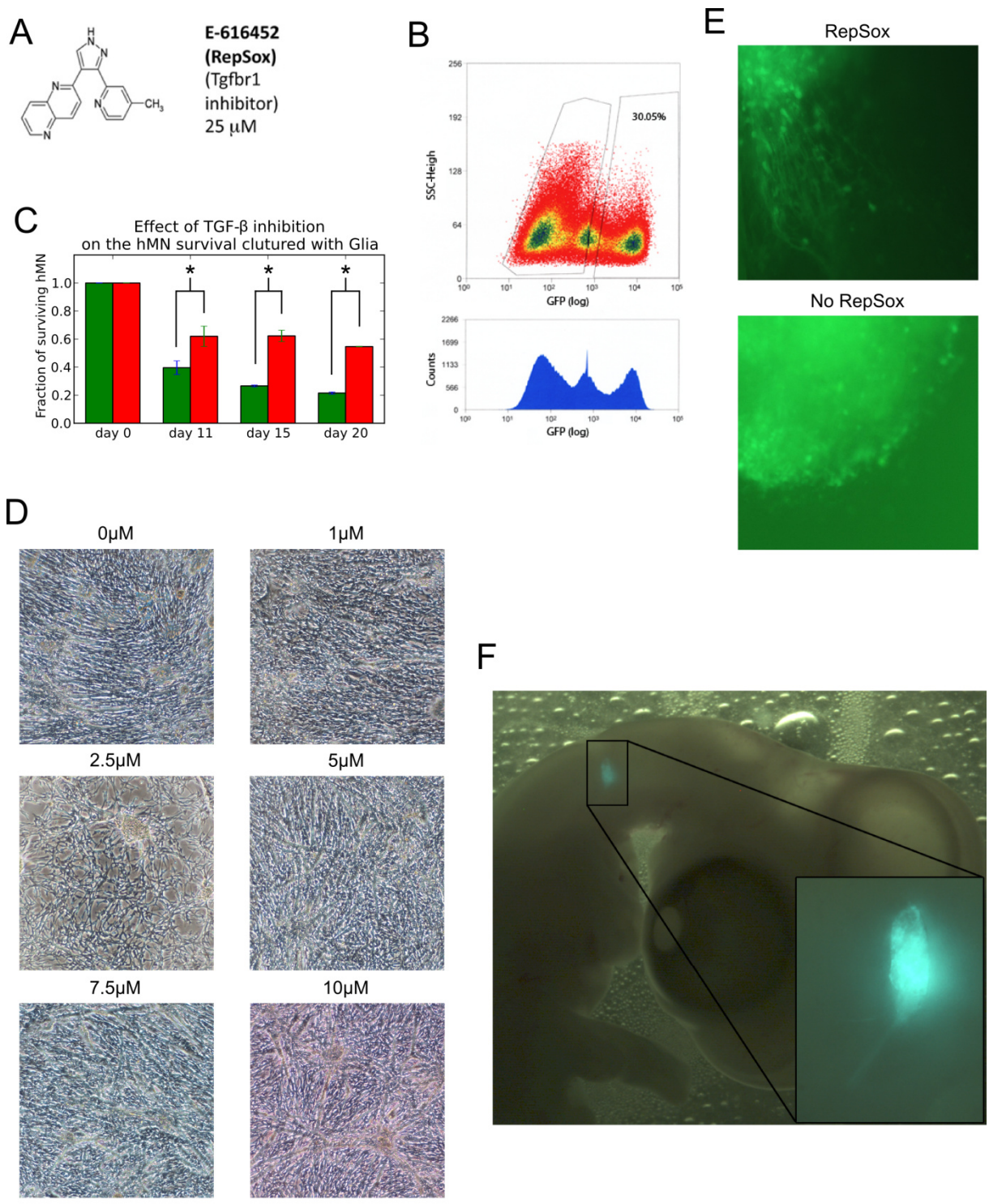



Figure 3: TGF-β inhibition on myotubes and hMNs
a) The RepSox molecule, a TGF-β inhibitor.
b) Flow cytometry sorting of dissociated HuES3 HB9::GFP Embryoid bodies. Only the high GFP and the highest low GFP populations were gated for the sorting.
c) Sorted HuES3 HB9::GFP MNs were plated on Poly-D-Lysin Laminin coated tissue culture dish with glia. Inhibition of TGF-β increase by 3-fold the survival. (n = 3, *: p-value < 0.01)
d) Effect of TGF-β inhibition on myotubes. There is a dose effect on the differentiation as the concentration of the inhibitor increase with 10µM as the more effective concentration (>10µM is toxic for the cells).
e) Representative picture of HuES3 HB9::GFP neuralized embryoid body at day 10. Inhibition of TGF-β increase drastically the axonal activity of the neuron.
f) hMNs embryoid body implanted in a chicken embryo (Stage HH16-18) and harvested 3 days later. It is possible to see some axons growing inside the embryo.

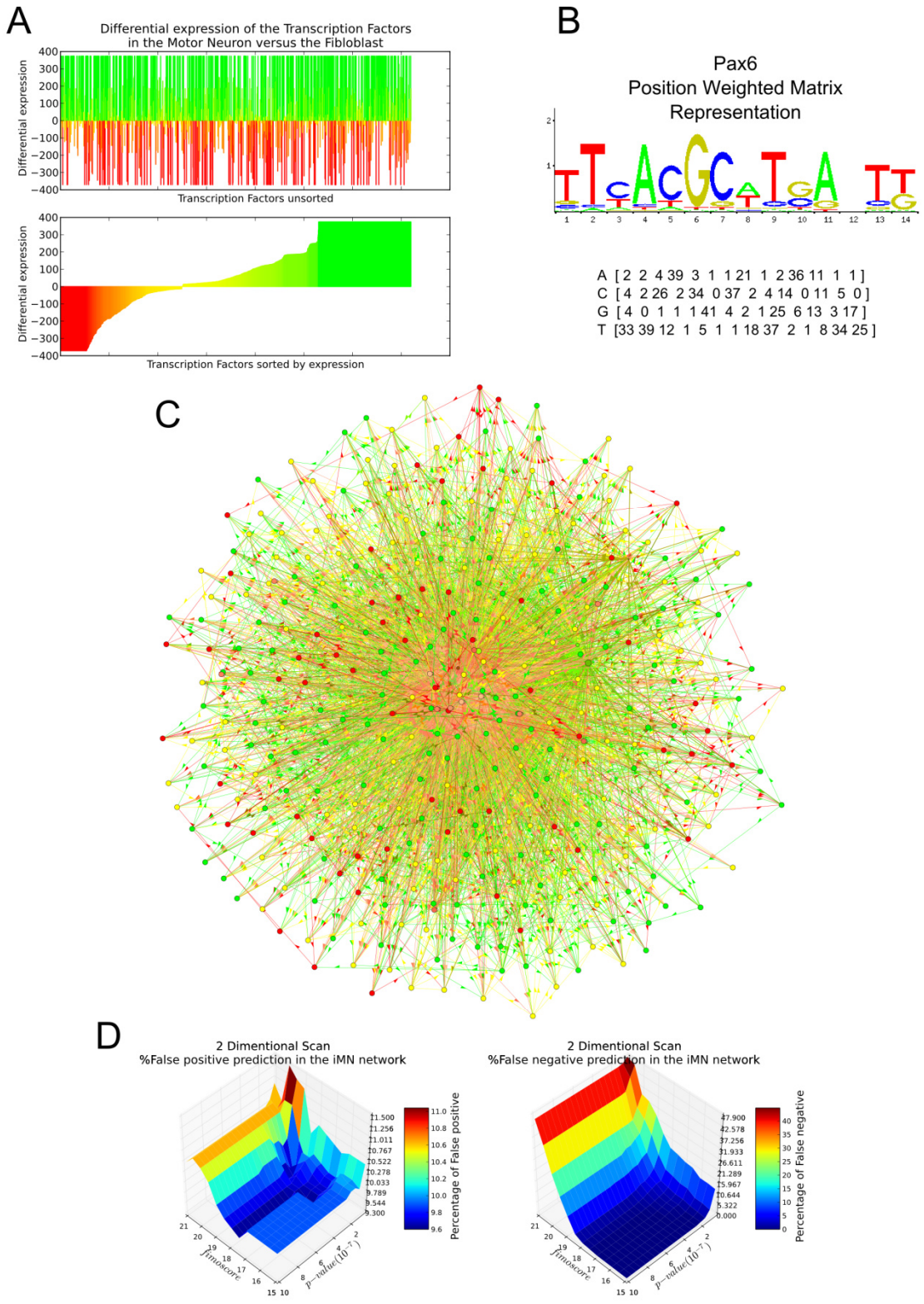



Figure 4: Gene regulatory network

a) Heatmap histogram of differential expressions of TFs. Microarray data from fibroblasts against MNs were used. Only differential expression scores with a p-value < 0.05 were selected.

b) Example the Pax6 Position Weighted Matrix. Each position of the sequence has a probability of being one of the 4 DNA bases. (Source : Jaspar database)

c) Representation of the iMN network starting with 20 factors and refined with the Microarray data. Only the Transcription Factors are shown here. (Red: Downregulated gene, Green: Upregulated gene, Yellow: No useable data in the Microarray)

d) Surface plot of a 2 Dimensional parameter scan. For the false positive, there is an area at fScore > 18 where the p-value isn't effecting the false positive percentage. The minimum is at p-value < 1.5 and fScore > 19. For the false negative, the more restricted it gets, the more false negatives.

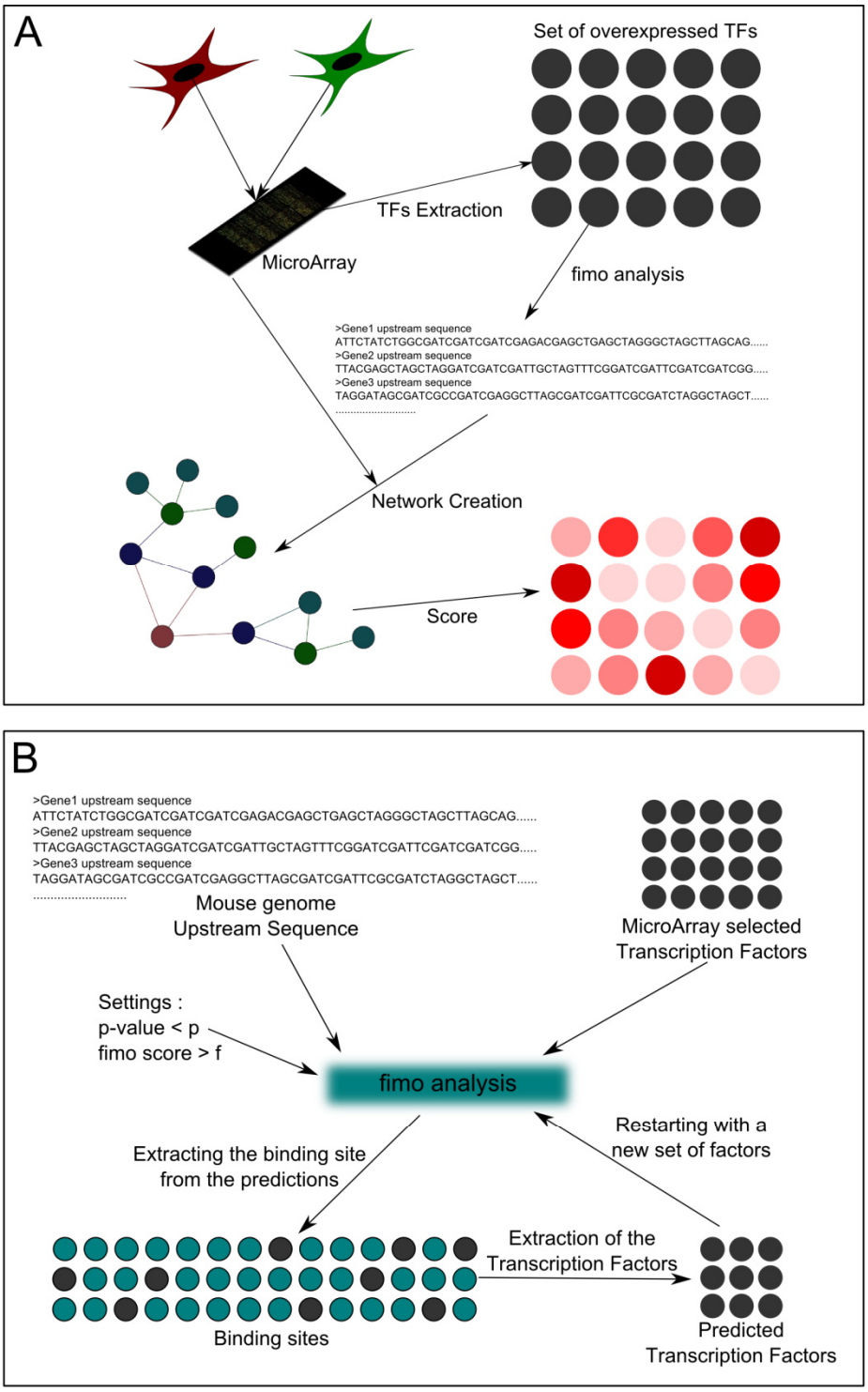
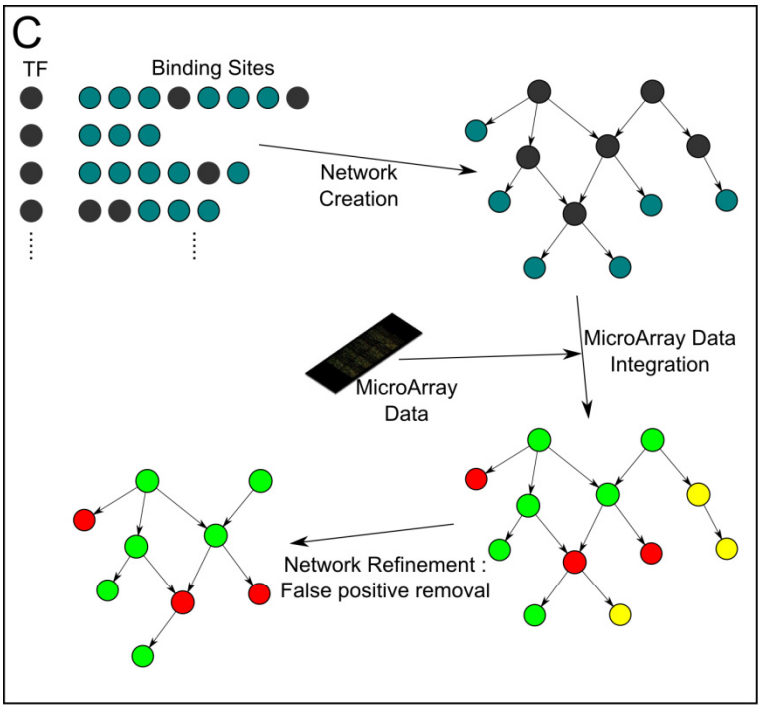
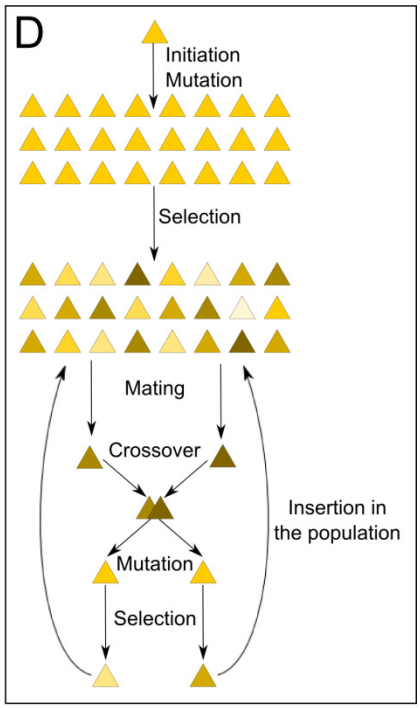
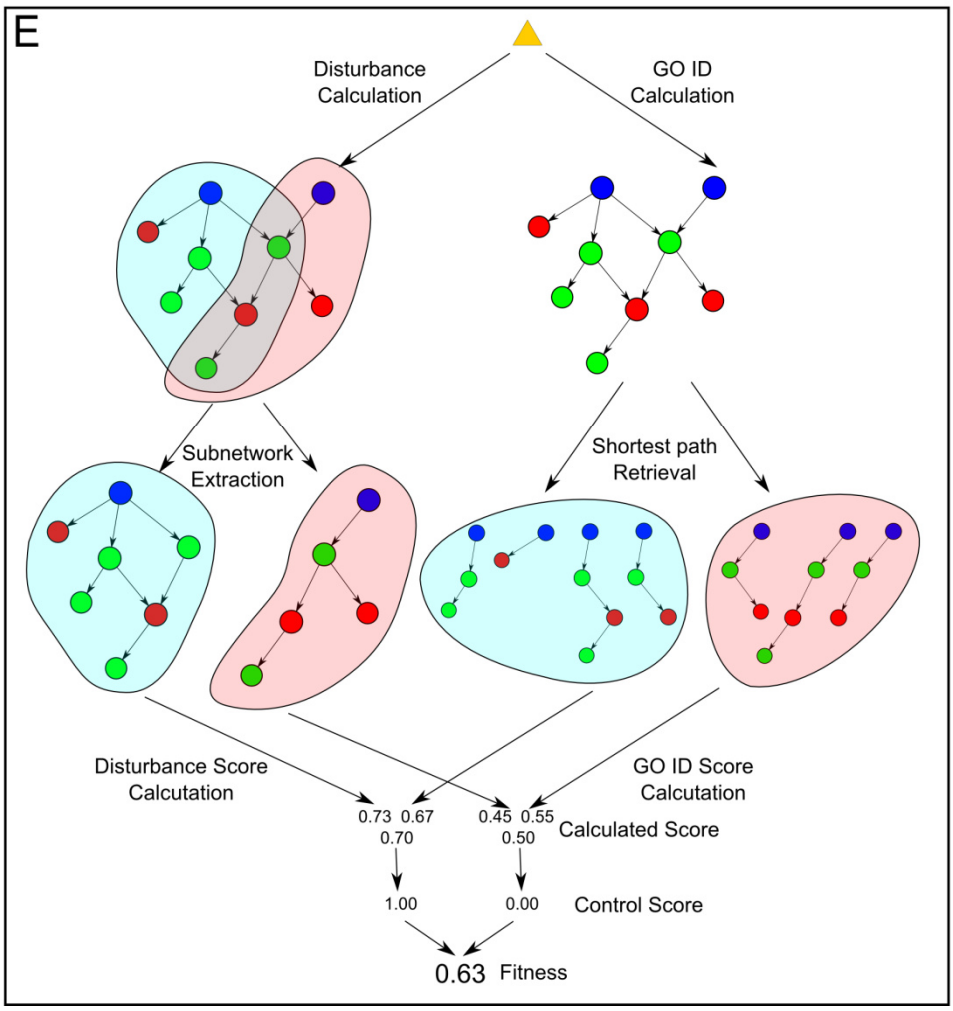



Figure 5: Concept and schematics of the transdifferentiation prediction algorithm
a) Global scheme of the algorithm, starting from a set of TFs, the gene regulation network is created, and the TFs are scored.
b) Representative scheme of the recursive fimo analysis, at each round the binding sites are stored with their corresponding TF. The algorithm loops 3 times. (Grey: TF, Blue : other gene)
c) Representative scheme of the network creation process and its Microarray refinement. Starting from the predicted binding sites, a graph is created, the expression levels are integrated in the graph, and all the genes which are not differentially expressed are removed from the network. (Green: upregulated, Red: downregulated, Yellow: no change in expression)
d) General scheme of the genetic algorithm. (Triangle: individual GO ID scoring table, Yellow gradient: represents the fitness of each individual)
e) Representative scheme of the selection process. Each individual is tested for its disturbance score and it's GO ID score for each TF. The two scores are merged and a correlation test is run versus the experimentally known value. The correlation coefficient is then used as a fitness.


C
TF
Binding Sites
Network
Creation
MicroArray
Data
MicroArray Data
Integration
Network Refinement :
False positive removal


D
Initiation
Mutation
Selection
Mating
Crossover
Mutation
Selection
Insertion in
the population


E
Disturbance
Calculation
GO ID
Calculation
Subnetwork
Extraction
Shortest path
Retrieval
Disturbance Score
Calcutation
GO ID Score
Calcutation
0.73 0.67
0.70
0.45 0.55
0.50
Calculated Score
1.00
0.00
Control Score
0.63
Fitness

# Bibliography

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, *25*(1), 25-9. Nature Publishing Group. doi: 10.1038/75556.

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *AAAI Press*, 28-36.

Boulland, J.-L., Halasi, G., Kasumacic, N., & Glover, J. C. (2010). Xenotransplantation of human stem cells into the chicken embryo. *Journal of visualized experiments : JoVE*, (41). doi: 10.3791/2071.

Boulting, G. L., Kiskinis, E., Croft, G. F., Amoroso, M. W., Oakley, D. H., Wainger, B. J., et al. (2011). A functionally characterized test set of human induced pluripotent stem cells. *Nature biotechnology*, *29*(3), 279-86. Nature Publishing Group. doi: 10.1038/nbt.1783.

Bryne, J. C., Valen, E., Tang, M.-H. E., Marstrand, T., Winther, O., Piedade, I. da, et al. (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic acids research*, *36*(Database issue), D102-6. Oxford University Press. doi: 10.1093/nar/gkm955.

Chiò, A., Schymick, J. C., Restagno, G., Scholz, S. W., Lombardo, F., Lai, S.-L., et al. (2009). A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Human molecular genetics*, *18*(8), 1524-32. doi: 10.1093/hmg/ddp059.

Dijkstra, E. W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik*, *1*, 269–271.

Dimos, J. T., Rodolfa, K. T., Niakan, K. K., Weisenthal, L. M., Mitsumoto, H., Chung, W., et al. (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science (New York, N.Y.)*, *321*(5893), 1218-21. American Association for the Advancement of Science. doi: 10.1126/science.1158799.

Efe, J. A., Hilcove, S., Kim, Janghwan, Zhou, H., Ouyang, K., Wang, G., et al. (2011). Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nature Cell Biology*, *13*(3), 215-222. Nature Publishing Group. doi: 10.1038/ncb2164.

Ge, X., McFarlane, C., Vajjala, A., Lokireddy, S., Ng, Z. H., Tan, C. K., et al. (2011). Smad3 signaling is required for satellite cell function and myogenic differentiation of myoblasts. *Cell research*. Nature Publishing Group. doi: 10.1038/cr.2011.72.

Gellibert, F., Woolven, J., Fouchet, M.-H., Mathews, N., Goodland, H., Lovegrove, V., et al. (2004). Identification of 1,5-naphthyridine derivatives as a novel series of potent and selective TGF-beta type I receptor inhibitors. *Journal of medicinal chemistry*, *47*(18), 4494-506. doi: 10.1021/jm0400247.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. *Addison Wesley*.

Ichida, J. K., Blanchard, J., Lam, K., Son, E. Y., Chung, J. E., Egli, D., et al. (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell stem cell*, *5*(5), 491-503. doi: 10.1016/j.stem.2009.09.012.

Karumbayaram, S., Kelly, T. K., Paucar, A. A., Roe, A. J. T., Umbach, J. A., Charles, A., et al. (2009). Human embryonic stem cell-derived motor neurons expressing SOD1 mutants exhibit typical signs of motor neuron degeneration linked to ALS. *Disease models & mechanisms*, *2*(3-4), 189-95. doi: 10.1242/dmm.002113.

Kim, Janghwan, Efe, J. A., Zhu, S., Talantova, M., Yuan, X., Wang, S., et al. (2011). Direct reprogramming of mouse fibroblasts to neural progenitors. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7838-43. doi: 10.1073/pnas.1103113108.

Klymkowsky, M. W., & Hanken, J. (1991). Whole-mount staining of Xenopus and other vertebrates. *Methods in cell biology*, *36*, 419-41. Retrieved June 12, 2011, from http://www.ncbi.nlm.nih.gov/pubmed/1725802.

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic acids research*, *34*(Database issue), D108-10. Oxford University Press. doi: 10.1093/nar/gkj143.

Mouse genome upstream sequences. (n.d.). . Retrieved from http://arep.med.harvard.edu/labgc/adnan/hsmmupstream/.

Orkin, S. H., Wang, J., Kim, J, Chu, J., Rao, S., Theunissen, T. W., et al. (2008). The Transcriptional Network Controlling Pluripotency in ES Cells. *Cold Spring Harbor symposia on quantitative biology*. doi: 10.1101/sqb.2008.72.001.

Pacak, C. A., & Cowan, D. B. (2009). Fabrication of myogenic engineered tissue constructs. *Journal of visualized experiments : JoVE*, (27). doi: 10.3791/1137.

Rafuse, V. F., & Landmesser, L. (1996). Contractile activity regulates isoform expression and polysialylation of NCAM in cultured myotubes: involvement of Ca2+ and protein kinase C. *The Journal of cell biology*, *132*(5), 969-83. Retrieved June 12, 2011, from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2120742&tool=pmcentrez&rendertype=abstract.

SAXEL, D. Y. & O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*.

Sha, Q., Zhang, Z., Schymick, J. C., Traynor, B. J., & Zhang, S. (2009). Genome-wide association reveals three SNPs associated with sporadic amyotrophic lateral sclerosis through a two-locus analysis. *BMC medical genetics*, *10*, 86. doi: 10.1186/1471-2350-10-86.

Shatunov, A., Mok, K., Newhouse, S., Weale, M. E., Smith, B., Vance, C., et al. (2010). Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet neurology*, *9*(10), 986-94. doi: 10.1016/S1474-4422(10)70197-6.

Shen, C. N., Horb, M. E., Slack, J. M. W., & Tosh, D. (2003). Transdifferentiation of pancreas to liver. *Mechanisms of development*, *120*(1), 107-16. Retrieved June 12, 2011, from http://www.ncbi.nlm.nih.gov/pubmed/12490301.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663-76. doi: 10.1016/j.cell.2006.07.024.

Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Südhof, T. C., & Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, *463*(7284), 1035-41. Nature Publishing Group. doi: 10.1038/nature08797.

Wasserman, W. W., & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics*, *5*(4), 276-287. Nature Publishing Group. doi: 10.1038/nrg1315.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520-62. doi: 10.1038/nature01262.